UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 13-23
Statistics and Econometrics Series 21
December 2013

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

A NEW GOODNESS-OF-FIT PROCESS FOR VARMA(p,q) MODELS: CONSTRUCTION AND EMPIRICAL PROPERTIES

Santiago Velilla[a] **and Huong Nguyen**[a]

Abstract

As an extension of the univariate technique in Ubierna and Velilla (2007), we present a goodness-of-fit process for *VARMA(p,q)* models in which the residuals of the fit are considered. We also formulate an explicit form of the asymptotic covariance function, as well as a suitable representation of the limit process. More importantly, we propose a new goodness-of-fit process based on a transformed correlation matrix sequence. The new goodness-of-fit process is proved to converge weakly to the Brownian bridge. Several simulations, comparisons, and examples are presented. These results illustrate the scope of both our theoretical findings and contributions. Our method is shown to be sensitive to detect lack of fit. Thus, it can be considered as a useful tool tool for identifying a proper time series model.

Keywords: Brownian bridge, error sample correlation matrix, goodness-of-fit process, model selection, VARMA(p,q) models, weak convergence.

---

[a]Universidad Carlos III de Madrid, Department of Statistics, Facultad de Ciencias Sociales y Jurídicas, Campus de Getafe, Madrid, Spain. E-mail addresses: santiago.velilla@uc3m.es (Santiago Velilla) and huong.nguyen@uc3m.es. (Huong Nguyen). Research partially supported by Grant ECO2011-25706 (Spain).

1 INTRODUCTION

1.1 Notations and definitions

Consider a causal and invertible $m-$variate autoregressive moving average *VARMA(p,q)* process

$$\boldsymbol{\Phi}(B)(\mathbf{X}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}(B)\boldsymbol{\varepsilon}_t \ , \tag{1}$$

where B is backward shift operator $B\mathbf{X}_t = \mathbf{X}_{t-1}$, $\boldsymbol{\mu}$ is the $m\times 1$ mean vector, and $\{\boldsymbol{\varepsilon}_t : t \in \mathbf{Z}\}$ is a zero mean white noise sequence $WN(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a $m \times m$ positive definite matrix. Additionally, $\boldsymbol{\Phi}(z) = \mathbf{I}_m - \boldsymbol{\Phi}_1 z - \cdots - \boldsymbol{\Phi}_p z^p$ and $\boldsymbol{\Theta}(z) = \mathbf{I}_m + \boldsymbol{\Theta}_1 z + \cdots + \boldsymbol{\Theta}_q z^q$ are matrix polynomials, where $\mathbf{I}_m$ is the $m \times m$ identity matrix, and $\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p, \boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_q$ are $m \times m$ real matrices such that the roots of the determinantal equations $|\boldsymbol{\Phi}(z)| = 0$ and $|\boldsymbol{\Theta}(z)| = 0$ all lie outside the unit circle. Additionally, the $m(p + q)$ roots are different from each other. It will be assumed that the conditions of Dunsmuir and Hannan (1976) hold, so that the model (1) is properly identified, and thus the terms of the operator $\boldsymbol{\Phi}^{-1}(B)\boldsymbol{\Theta}(B) = \sum_{j=0}^{\infty} \boldsymbol{\Omega}_j B^j$ are uniquely defined. In particular, both $\boldsymbol{\Phi}_p$ and $\boldsymbol{\Theta}_q$ are non-null matrices, and that the identifiability condition of Hannan (1969), $r(\boldsymbol{\Phi}_p, \boldsymbol{\Theta}_q) = m$, holds. In what follows, it will be convenient to put $P = \max(p, q)$, and to define the $m \times mp$ matrix $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p)$, the $m \times mq$ matrix $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_q)$, and the $m^2(p + q) \times 1$ vector of parameters $\boldsymbol{\Lambda} = \mathrm{vec}(\boldsymbol{\Phi}, \boldsymbol{\Theta})$.

Given n observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from model (1), the mean vector $\boldsymbol{\mu}$ can be estimated by the sample mean $\overline{\mathbf{X}}_n = n^{-1} \sum_{t=1}^{n} \mathbf{X}_t$. In general, finding the ML estimates $(\widehat{\boldsymbol{\Phi}}_n, \widehat{\boldsymbol{\Theta}}_n, \widehat{\boldsymbol{\Sigma}}_n)$ of $(\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\Sigma})$ is a complex nonlinear optimization problem that must be solved using an adequate efficient algorithm (see e.g. Lütkepohl, 2005, section 12.3.5). According to Lütkepohl (2005, p. 408), the vector of ML estimators $\widehat{\boldsymbol{\Lambda}} = \mathrm{vec}(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Theta}})$ is consistent and asymptotically normal for $\boldsymbol{\Lambda} = \mathrm{vec}(\boldsymbol{\Phi}, \boldsymbol{\Theta})$. In particular,

$$\sqrt{n}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}) \xrightarrow{D} N_{m^2(p+q)}[\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\Lambda})] \ , \tag{2}$$

where $\mathbf{I}(\boldsymbol{\Lambda})$ is the $m^2(p + q) \times m^2(p + q)$ information matrix.

The residual vectors $\widehat{\boldsymbol{\varepsilon}}_t$, $t = 1, \ldots, n$, are defined recursively in the form

$$\widehat{\boldsymbol{\varepsilon}}_t = \boldsymbol{\varepsilon}_t(\widehat{\boldsymbol{\Theta}}, \widehat{\boldsymbol{\Phi}}, \overline{\mathbf{X}}_n) = (\mathbf{X}_t - \overline{\mathbf{X}}_n) - \sum_{i=1}^{p} \widehat{\boldsymbol{\Phi}}_i(\mathbf{X}_{t-i} - \overline{\mathbf{X}}_n) - \sum_{j=1}^{q} \widehat{\boldsymbol{\Theta}}_j \widehat{\boldsymbol{\varepsilon}}_{t-j} \ , \quad t = 1, \ldots, n \ , \tag{3}$$

with the usual conditions $\mathbf{X}_t - \overline{\mathbf{X}}_n \equiv \mathbf{0} \equiv \widehat{\boldsymbol{\varepsilon}}_t$, for $t \leq 0$. In practice, only residual vectors for $t > P = \max(p, q)$ are considered.

1.2 Motivation

Ubierna and Velilla (2007) considered a goodness-of-fit process in *ARMA(p,q)* models based on a modified autocorrelation sequence. The basic idea is to start from the process $\{\widehat{W}_n(u) : 0 \leq u \leq 1\}$, where

$$\widehat{W}_n(u) = \sqrt{\frac{n}{2}}\,[\widehat{F}_n(\pi u) - F_0(\pi u)] = \frac{\sqrt{2}}{\pi}\,\sqrt{n}\sum_{k=1}^{n-(P+1)} \widehat{r}_k\,\frac{\sin(k\pi u)}{k} \tag{4}$$

is a random element in $C[0,1]$, the space of continuous functions in $[0,1]$. According to Durbin (1975, sec. 2), $\{\widehat{W}_n(u) : 0 \leq u \leq 1\}$ converges weakly, as $n \to \infty$, to a centered Gaussian process $\{G(u) : 0 \leq u \leq 1\}$ with covariance function

$$\gamma(u,v) = [\min(u,v) - uv] - (2\pi^2)^{-1} g(\pi u)' \mathbf{I}^{-1}(\boldsymbol{\phi}, \boldsymbol{\theta}) g(\pi v)\ , \qquad 0 \leq u, v \leq 1\ , \tag{5}$$

where $\mathbf{I}(\boldsymbol{\phi}\ , \boldsymbol{\theta})$ is the $(p+q)\times(p+q)$ information matrix, and $g(\pi u) = \int_0^{\pi u}[\partial \log f(t)/\partial(\boldsymbol{\phi}\ , \boldsymbol{\theta})]dt$ is a $(p+q)\times 1$ vector that depends on the normalized spectral density function $f(.)$ of an *ARMA(p,q)* process.

By the continuous mapping theorem, the asymptotic distribution of any continuous functional $H[\widehat{W}_n(u)]$ of the process (4) is given by $H[G(u)]$. Accordingly, since the covariance function of the limit process depends on the unknown parameters, using $H[\widehat{W}_n(u)]$ for goodness-of-fit purposes is not feasible. The performance of the new method in Ubierna and Velilla (2007) is shown to be better than standard methods of goodness-of-fit. As an extension of that approach, we study a goodness-of-fit process $\{W_n^m(u) : 0 \leq u \leq 1\}$ for *VARMA(p,q)* models, where

$$W_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-1} \frac{\mathrm{tr}(\mathbf{R}_k)}{\sqrt{m}}\,\frac{\sin(k\pi u)}{k} \tag{6}$$

depends on the error correlation matrix $\mathbf{R}_k = \mathbf{C}_k'\mathbf{C}_0^{-1}$, introduced by Chitturi (1974). Since $W_n^m(u)$ depends on the unobservable errors $\{\boldsymbol{\varepsilon}_t : 1 \leq t \leq n\}$, in practice the residual version $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ is considered, where

$$\widehat{W}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\,\frac{\sin(k\pi u)}{k}\ , \tag{7}$$

and $\widehat{\mathbf{R}}_k = \widehat{\mathbf{C}}_k \widehat{\mathbf{C}}_0^{-1}$ is the correlation matrix of the residuals $\widehat{\boldsymbol{\varepsilon}}_t$.

The rest of this working paper is organized as follows. Section 2 presents some auxiliary results. Section 3 proposes a new goodness-of-fit process based on a transformed correlation matrix sequence. Properties of the associated transformation technique are considered. The new goodness-of-fit process is proved to converge weakly to the Brownian bridge. Section 4 contains an empirical exploration, for several *VARMA(p,q)* models, of its behavior in both size and power. Section 5 gives some final conclusions.

2 AUXILIARY RESULTS

Let $\mathbf{C}_k = n^{-1} \sum_{t=1}^{n-k} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_{t+k}, 0 \leq k \leq n-1$, denote the kth $m \times m$ sample error covariance matrix at lag k, and $\widehat{\mathbf{C}}_k = n^{-1} \sum_{t>P}^{n-k} \widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}'_{t+k}, 0 \leq k \leq n-(P+1)$, the associated $m \times m$ residual version. For convenience, write $\widehat{\boldsymbol{\Sigma}} = \widehat{\mathbf{C}}_0$. Consider the $km^2 \times 1$ random vectors $\widehat{\mathbf{H}}_k = [\mathrm{vec}(\widehat{\mathbf{C}}'_1), \ldots, \mathrm{vec}(\widehat{\mathbf{C}}'_k)]'$, and $\mathbf{H}_k = [\mathrm{vec}(\mathbf{C}'_1), \ldots, \mathrm{vec}(\mathbf{C}'_k)]'$.

Define on the other hand the $m \times m$ coefficients of the series expansions $\boldsymbol{\Phi}^{-1}(z)\boldsymbol{\Theta}(z) = \sum_{j=0}^{\infty} \boldsymbol{\Omega}_j z^j$, and $\boldsymbol{\Theta}^{-1}(z) = \sum_{j=0}^{\infty} \mathbf{L}_j z^j$, where $\boldsymbol{\Omega}_0 = \mathbf{L}_0 = \mathbf{I}_m$. Define also the collection of matrices $\mathbf{G}_k = \sum_{j=0}^{k} (\boldsymbol{\Sigma}\boldsymbol{\Omega}'_j \otimes \mathbf{L}_{k-j})$, and $\mathbf{F}_k = \boldsymbol{\Sigma} \otimes \mathbf{L}_k$, $k \geq 0$. Let $\mathbf{Z}_k = (\mathbf{X}_k, \mathbf{Y}_k)$ denote a $km^2 \times m^2(p+q)$ matrix which is such that

$$\mathbf{X}_k = \begin{pmatrix} \mathbf{G}_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{G}_1 & \mathbf{G}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{G}_2 & \mathbf{G}_1 & \mathbf{G}_0 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{G}_{k-1} & \mathbf{G}_{k-2} & \mathbf{G}_{k-3} & \cdots & \mathbf{G}_{k-p} \end{pmatrix},$$

and

$$\mathbf{Y}_k = \begin{pmatrix} \mathbf{F}_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F}_1 & \mathbf{F}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F}_2 & \mathbf{F}_1 & \mathbf{F}_0 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{F}_{k-1} & \mathbf{F}_{k-2} & \mathbf{F}_{k-3} & \cdots & \mathbf{F}_{k-q} \end{pmatrix}.$$

Following Lemma 2 (Hosking, 1980, p. 603), it follows that, for each $k \geq 1$,

$$\widehat{\mathbf{H}}_k = \mathbf{H}_k - \mathbf{Z}_k \mathrm{vec}[(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Theta}}) - (\boldsymbol{\Phi}, \boldsymbol{\Theta})] + O_P(\frac{1}{n}) \,. \tag{8}$$

On the other hand, according to Hosking (1980, Appendix, p.607), the orthogonality condition $\mathbf{Z}_k'\mathcal{W}^{-1}\widehat{\mathbf{H}}_k = O_P(1/n)$ holds, where $\mathcal{W} = \mathbf{I}_k \otimes \mathbf{\Sigma} \otimes \mathbf{\Sigma}$.

2.1 Representation of the limit process

Consider the $m^2 \times m^2$ function $\mathbf{P}_k(u) = \sum_{j=1}^{\infty}(\delta_{jk}\mathbf{I}_{m^2} - \mathbf{P}_{jk})\sin(j\pi u)/j$, $0 \leq u \leq 1$, where δ_{jk} is Dirac's delta. On the other hand, $\mathbf{P}_{jk} = \mathbf{\Xi}_j'\mathbf{I}^{-1}(\mathbf{\Lambda})\mathbf{\Xi}_k$, $j, k \geq 1$, where $\mathbf{I}(\mathbf{\Lambda})$ is the $m^2(p+q) \times m^2(p+q)$ information matrix, and $\mathbf{\Xi}_j'$ the jth $m^2 \times m^2(p+q)$ row block of the matrix $\mathcal{W}^{-1/2}\mathbf{Z}_M$.

The following Gaussian process $\{G^m(u) : 0 \leq u \leq 1\}$ is proposed as a limit candidate for $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ in (7),

$$G^m(u) = \frac{\sqrt{2}}{\pi}\sum_{k=1}^{\infty}\mathbf{a}'\mathbf{P}_k(u)\mathbf{V}_k \ , \tag{9}$$

where the $\{\mathbf{V}_k : k \geq 1\}$ are i.i.d. $\mathbf{N}_{m^2}(\mathbf{0}, \mathbf{I}_{m^2})$ random vectors. It can be shown that the limit covariance function $\gamma^m(u, v)$ of the process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ can be written

$$\gamma^m(u,v) = \frac{2}{\pi^2}[\sum_{k=1}^{\infty}\frac{\sin(k\pi u)\sin(k\pi v)}{k^2} - \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}(\mathbf{a}'\mathbf{P}_{jk}\mathbf{a})\frac{\sin(j\pi u)}{j}\frac{\sin(k\pi v)}{k}] \ . \tag{10}$$

The first summand in (10) is the covariance function of the Brownian bridge $\{B(u) : 0 \leq u \leq 1\}$. The second depends on the covariance matrix $\mathbf{\Sigma}$ of the errors through the information matrix $\mathbf{I}(\mathbf{\Lambda})$.

2.2 Weak convergence of the residual process

This section formalizes the weak convergence of the residual process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ to the Gaussian random function of (9).

Theorem 2.1 Assume that the errors of model (1) have finite eighth order moments $\mathrm{E}[\|\boldsymbol{\varepsilon}_t\|^8] < +\infty$. Then, as $n \to \infty$,

$$\widehat{W}_n^m(u) \to_\omega G^m(u) \ , \tag{11}$$

where $\{G^m(u) : 0 \leq u \leq 1\}$ is the zero mean Gaussian process of (9) with covariance function $\gamma^m(u, v), 0 \leq u, v \leq 1$, given in (10).

Proof. The proof is based on Velilla and Nguyen (2011, Lemma 3.1). Details are omitted for reasons of conciseness.■

3 A MODIFIED MULTIVARIATE GOODNESS-OF-FIT PROCESS

By the continuous mapping theorem, the asymptotic distribution of any continuous functional $H[\widehat{W}_n^m(u)]$ of the process (7) is given by $H[G^m(u)]$. As seen above, $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ has a complex asymptotic distribution that depends on the unknown parameters $\mathbf{\Lambda}$ and $\mathbf{\Sigma}$. Therefore, assessing the significance of an observed value of $H[\widehat{W}_n^m(u)]$ with $H[G^m(u)]$ is not feasible. As a possible solution, a goodness-of-fit process based on a transformed correlation matrix sequence is proposed next.

3.1 Motivation and definitions

For *ARMA(p,q)* models, Ubierna and Velilla (2007) consider a modified goodness-of-fit process $\{\widehat{Z}_n(u) : 0 \leq u \leq 1\}$, where

$$\widehat{Z}_n(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=p+q+1}^{n-(P+1)} \widehat{s}_k \frac{\sin(K\pi u)}{K} \; ; \tag{12}$$

$K = k - (p+q)$; and $\{\widehat{s}_k : p+q+1 \leq k \leq n-(P+1)\}$ is a transformed residual autocorrelation sequence. The goal of this section is to propose an adequate generalization of the process of (12) for *VARMA(p,q)* models. A new goodness-of-fit process $\{\widehat{Z}_n^m(u) : 0 \leq u \leq 1\}$ is introduced, where

$$\widehat{Z}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=p+q+1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}} \frac{\sin(K\pi u)}{K} \tag{13}$$

depends on a modified residual matrix autocorrelation sequence $\{\widehat{\mathbf{S}}_k\}$.

Specifically, the $m \times m$ matrices $\{\widehat{\mathbf{S}}_k\}$ are constructed so that

$$\begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{S}}_{p+q+1}) \\ \mathrm{vec}(\widehat{\mathbf{S}}_{p+q+2}) \\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{S}}_M) \end{pmatrix} = \widehat{\mathbf{\Psi}}'_M \begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_1\widehat{\mathbf{\Sigma}}^{-1/2}) \\ \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_2\widehat{\mathbf{\Sigma}}^{-1/2}) \\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}'_M\widehat{\mathbf{\Sigma}}^{-1/2}) \end{pmatrix} , \tag{14}$$

where $\widehat{\mathbf{\Psi}}_M$ is the estimated version of a $Mm^2 \times m^2[M-(p+q)]$ matrix $\mathbf{\Psi}_M$ such that

$$\mathbf{\Psi}'_M \mathbf{\Psi}_M = \mathbf{I}_{m^2[M-(p+q)]} \quad , \quad \mathbf{\Psi}'_M \mathcal{W}^{-1/2} \mathbf{Z}_M = \mathbf{0} \ . \tag{15}$$

From Hosking (1980),

$$\sqrt{n} \begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2} \widehat{\mathbf{C}}'_1 \widehat{\mathbf{\Sigma}}^{-1/2}) \\ \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2} \widehat{\mathbf{C}}'_2 \widehat{\mathbf{\Sigma}}^{-1/2}) \\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2} \widehat{\mathbf{C}}'_M \widehat{\mathbf{\Sigma}}^{-1/2}) \end{pmatrix} \overset{D}{\cong} (\mathbf{I}_{Mm^2} - \mathbf{P}_M) N_{Mm^2}(\mathbf{0}, \mathbf{I}_{Mm^2}) \ . \tag{16}$$

Combining (16)−(14)−(15) leads to

$$\sqrt{n} \begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{S}}_{p+q+1}) \\ \mathrm{vec}(\widehat{\mathbf{S}}_{p+q+2}) \\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{S}}_M) \end{pmatrix} \overset{D}{\cong}$$

$$\overset{D}{\cong} \mathbf{\Psi}'_M (\mathbf{I}_{Mm^2} - \mathbf{P}_M) N_{Mm^2}(\mathbf{0}, \mathbf{I}_{Mm^2}) = N_{m^2[M-(p+q)]}(\mathbf{0}, \mathbf{I}_{m^2[M-(p+q)]}) \ . \tag{17}$$

Consider now the sequence of modified adjusted residual traces

$$\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m} = \mathbf{a}'_m \mathrm{vec}(\widehat{\mathbf{S}}_k) \ , \qquad p+q+1 \le k \le n-(P+1) \ , \tag{18}$$

where $\mathbf{a}_m = \mathrm{vec}(\mathbf{I}_m)/\sqrt{m}$ is a $m^2 \times 1$ unit vector. Since $[\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m][\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}_m] = \mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m \mathbf{a}_m = \mathbf{I}_{M-(p+q)} \otimes 1 = \mathbf{I}_{M-(p+q)}$, from (17) it follows that

$$\sqrt{n}\,[\frac{1}{\sqrt{m}} \begin{pmatrix} \mathrm{tr}(\widehat{\mathbf{S}}_{p+q+1}) \\ \mathrm{tr}(\widehat{\mathbf{S}}_{p+q+2}) \\ \vdots \\ \mathrm{tr}(\widehat{\mathbf{S}}_M) \end{pmatrix}] =$$

$$= \sqrt{n}\,[\mathbf{I}_{M-(p+q)} \otimes \mathbf{a}'_m] \begin{pmatrix} \mathrm{vec}(\widehat{\mathbf{S}}_{p+q+1}) \\ \mathrm{vec}(\widehat{\mathbf{S}}_{p+q+2}) \\ \vdots \\ \mathrm{vec}(\widehat{\mathbf{S}}_M) \end{pmatrix} \overset{D}{\cong} N_{M-(p+q)}[\mathbf{0}, \mathbf{I}_{M-(p+q)}] \ . \tag{19}$$

As a final conclusion from (19),

$$\sqrt{n}\,[\frac{1}{\sqrt{m}}\begin{pmatrix}\mathrm{tr}(\widehat{\mathbf{S}}_{p+q+1})\\ \mathrm{tr}(\widehat{\mathbf{S}}_{p+q+2})\\ \vdots\\ \mathrm{tr}(\widehat{\mathbf{S}}_{M})\end{pmatrix}] \stackrel{D}{\cong} \begin{pmatrix} v_1\\ v_2\\ \vdots\\ v_{M-(p+q)}\end{pmatrix}\ , \tag{20}$$

where $\{v_k : k \geq 1\}$ is a sequence of i.i.d. $N(0,1)$ random variables. It can be then conjectured that (13) will converge to the Brownian Bridge, since its partial sums will approach to those of this last process.

3.2 Construction and properties of the sequence of transformation matrices

The $Mm^2 \times m^2[M-(p+q)]$ matrix $\mathbf{\Psi}_M$ of (15) is constructed by considering the collection of $Mm^2 \times m^2$ matrices

$$\mathbf{\Gamma}_k^* = \left(\frac{\mathbf{\Gamma}_k}{\mathbf{0}_{(M-k)m^2\times m^2}}\right)\ , \quad k = p+q+1,\ldots,M\ , \tag{21}$$

where $\mathbf{0}_{(M-k)m^2\times m^2}$ is a zero matrix, and defining on the other hand

$$\mathbf{\Psi}_M = (\mathbf{\Gamma}_{p+q+1}^* \,|\, \mathbf{\Gamma}_{p+q+2}^* \,|\cdots|\, \mathbf{\Gamma}_M^*)\ , \tag{22}$$

where, for each $p+q+1 \leq k \leq n-(P+1)$, $\mathbf{\Gamma}_k = \mathbf{\Delta}_k(\mathbf{\Delta}_k'\mathbf{\Delta}_k)^{-1/2}$ and

$$\mathbf{\Delta}_k = \begin{pmatrix} -\mathcal{W}^{-1/2}\mathbf{Z}_{k-1}(\mathbf{Z}_{k-1}'\mathcal{W}^{-1}\mathbf{Z}_{k-1})^{-1}\mathbf{\Xi}_k\\ \mathbf{I}_{m^2}\end{pmatrix} \tag{23}$$

is a matrix of $km^2 \times m^2$.

According to he above expressions, the kth modified residual correlation $m\times m$ matrix $\widehat{\mathbf{S}}_k$ is such that

$$\mathrm{vec}(\widehat{\mathbf{S}}_k) = \widehat{\mathbf{\Gamma}}_k'\begin{pmatrix}\mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}_1'\widehat{\mathbf{\Sigma}}^{-1/2})\\ \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}_2'\widehat{\mathbf{\Sigma}}^{-1/2})\\ \vdots\\ \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}_k'\widehat{\mathbf{\Sigma}}^{-1/2})\end{pmatrix}\ , \quad p+q+1 \leq k \leq n-(P+1)\ , \tag{24}$$

where $\widehat{\mathbf{\Gamma}}_k = \widehat{\mathbf{\Delta}}_k(\widehat{\mathbf{\Delta}}_k'\widehat{\mathbf{\Delta}}_k)^{-1/2}$ is an empirical version of $\mathbf{\Gamma}_k = \mathbf{\Delta}_k(\mathbf{\Delta}_k'\mathbf{\Delta}_k)^{-1/2}$, that is constructed from suitable ML estimators $(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Theta}}, \widehat{\mathbf{\Sigma}})$ of the parameters $(\mathbf{\Phi}, \mathbf{\Theta}, \mathbf{\Sigma})$ of model (1). From expression (24), it can be written

$$\mathrm{vec}(\widehat{\mathbf{S}}_k) = [\mathbf{I}_{m^2} + \widehat{\mathbf{\Xi}}_k'(\widehat{\mathbf{Z}}_{k-1}'\widehat{\mathcal{W}}^{-1}\widehat{\mathbf{Z}}_{k-1})^{-1}\widehat{\mathbf{\Xi}}_k]^{-1/2}$$

$$[\,\mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}_k'\widehat{\mathbf{\Sigma}}^{-1/2}) - \widehat{\mathbf{\Xi}}_k'(\widehat{\mathbf{Z}}_{k-1}'\widehat{\mathcal{W}}^{-1}\widehat{\mathbf{Z}}_{k-1})^{-1}\sum_{j=1}^{k-1}\widehat{\mathbf{\Xi}}_j\mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}_j'\widehat{\mathbf{\Sigma}}^{-1/2})\,]\;, \qquad (25)$$

$p+q+1 \leq k \leq n-(P+1)$. In practice, the blocks $\mathbf{\Xi}_j'$ are exponentially bounded, so for n and k large enough the numerical approximation $\mathrm{vec}(\widehat{\mathbf{S}}_k) \cong \mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}_k'\widehat{\mathbf{\Sigma}}^{-1/2})$ is obtained. By considering the $m^2 \times 1$ unit vector $\mathbf{a}_m = \mathrm{vec}(\mathbf{I}_m)/\sqrt{m}$, it follows finally that

$$\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m} = \mathbf{a}_m'\mathrm{vec}(\widehat{\mathbf{S}}_k) \cong \mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}\;. \qquad (26)$$

From (26), the modified adjusted traces $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ will tend to be similar to the original ones $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ for both n and k large. Relation (26) is investigated later in simulations.

3.3 Convergence to the Brownian bridge

Theorem 3.1 studies the limit properties of the modified process $\{\widehat{Z}_n^m(u) : 0 \leq u \leq 1\}$,

$$\widehat{Z}_n^m(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\sum_{k=p+q+1}^{n-(P+1)}\frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}\frac{\sin(K\pi u)}{K}\;,$$

where $K = k-(p+q)$.

Theorem 3.1 Under the same assumptions for the errors of model (1) than in Theorem 2.1, as $n \to \infty$,

$$\widehat{Z}_n^m(u) \to_w B(u)\;. \qquad (27)$$

Proof. The technique of proof is again based on Lemma 3.1 (Velilla and Nguyen, 2011). Details are omitted for reasons of conciseness. ■

4 EMPIRICAL WORK

Simulation techniques are used to investigate the behavior of the technique presented here. For the case of pure autoregressive *VAR(p)* models, the Yule-Walker estimators $\widehat{\mathbf{\Phi}}_i$,

$i = 1, \ldots, p$, are used. For *VARMA(p,q)* processes, the algorithm is the ML estimation scheme suggested by Lütkepohl (2005, sec.12.3). In all the simulations presented below, the mean vector $\boldsymbol{\mu} = \mathrm{E}(\mathbf{X}_t)$ is always taken to be equal to $\mathbf{0}$.

4.1 Examples of *VARMA(p,q)* processes

4.1.1 The *VAR(1)* model

Consider the *VAR(1)* model for $m = 2$,

$$\mathbf{X}_t = \boldsymbol{\Phi}_1 \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t \ , \tag{28}$$

where

$$\boldsymbol{\Phi}_1 = \begin{pmatrix} 0.2802 & 0.2680 \\ -0.0183 & 0.3152 \end{pmatrix} \ . \tag{29}$$

The matrix of (29) is obtained by taking eigenvalues $\delta_j = 0.2977 \pm 0.0678\, i$, $j = 1, 2$, so that $|\delta_1| = |\delta_2| = 0.3053 < 1$. The associated eigenvectors are selected in the form $\boldsymbol{\gamma}_1 = (2.7071, 0.1768 + 0.6847\, i)'$, and $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (2.7071, 0.1768 - 0.6847\, i)'$. Thus, the array of (29) follows from the identity $\boldsymbol{\Phi}_1 = \mathbf{C}\mathbf{D}\mathbf{C}^{-1}$, where $\mathbf{C} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ and $\mathbf{D} = \mathrm{diag}(\delta_1, \delta_2)$. The covariance matrix of the errors $\boldsymbol{\varepsilon}_t$ in (28) will be given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{pmatrix} \ . \tag{30}$$

In this case, the jth row-block of $\mathcal{W}^{-1/2}\mathbf{Z}_M$ is of the form

$$\boldsymbol{\Xi}_j' = \boldsymbol{\Sigma}^{1/2}(\boldsymbol{\Phi}_1^{j-1})' \otimes \boldsymbol{\Sigma}^{-1/2} \ , \quad j \geq 1 \ .$$

For M large enough, the information matrix can be approximated by the sum

$$\mathbf{Z}_M' \mathcal{W}^{-1} \mathbf{Z}_M = \sum_{j=1}^{M} \boldsymbol{\Xi}_j \boldsymbol{\Xi}_j' = [\sum_{j=0}^{M-1} \boldsymbol{\Phi}_1^j \boldsymbol{\Sigma} (\boldsymbol{\Phi}_1^j)'] \otimes \boldsymbol{\Sigma}^{-1} \ ,$$

where by convention $\boldsymbol{\Phi}_1^0 = \mathbf{I}_m$. It is found numerically that $\|\boldsymbol{\Xi}_j\| < 10^{-15}$ for $j > 6$. Thus, only $M = 6$ terms are really needed in the sum above. This produces the following information matrix:

$$\mathbf{I}[\mathrm{vec}(\mathbf{\Phi}_1)] = \begin{pmatrix} 1.3628 & -0.4088 & 0.4640 & -0.1392 \\ -0.4088 & 1.3628 & -0.1392 & 0.4640 \\ 0.4640 & -0.1392 & 1.2147 & -0.3644 \\ -0.1392 & 0.4640 & -0.3644 & 1.2147 \end{pmatrix} . \tag{31}$$

The structure of (31) resembles to that of the matrix

$$\mathbf{\Sigma} \otimes \mathbf{\Sigma}^{-1} = \begin{pmatrix} 1.0989 & -0.3297 & 0.3297 & -0.0989 \\ -0.3297 & 1.0989 & -0.0989 & 0.3297 \\ 0.3297 & -0.0989 & 1.0989 & -0.3297 \\ -0.0989 & 0.3297 & -0.3297 & 1.0989 \end{pmatrix} , \tag{32}$$

where

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} 1.0989 & -0.3297 \\ -0.3297 & 1.0989 \end{pmatrix} . \tag{33}$$

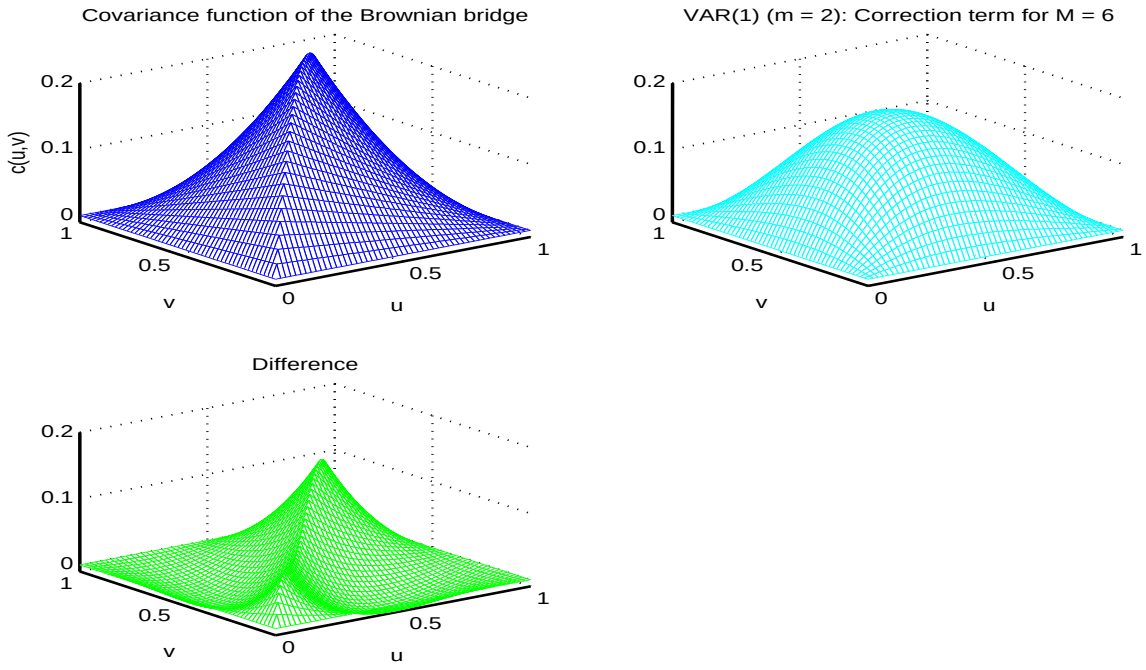

FIG. 1: Covariance function for the limit of the residual process of (7) under the bivariate *VAR(1)* model (28)

It is of interest now to analyze the magnitude of the coefficients $\mathbf{a}_2'\mathbf{P}_{jk}\mathbf{a}_2$ that appear in the second summand of the covariance function,

$$\gamma^2(u,v) = [\min(u,v) - uv] - \frac{2}{\pi^2}\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}(\mathbf{a}_2'\mathbf{P}_{jk}\mathbf{a}_2)\frac{\sin(j\pi u)}{j}\frac{\sin(k\pi v)}{k}\ ,$$

$0 \leq u, v \leq 1$, where $\mathbf{P}_{jk} = \mathbf{\Xi}_j'\mathbf{I}^{-1}[\mathrm{vec}(\mathbf{\Phi}_1)]\mathbf{\Xi}_k$, $j, k \geq 1$, and $\mathbf{a}_2 = \mathrm{vec}(\mathbf{I}_2)/\sqrt{2} = (1,0,0,1)'/\sqrt{2}$. As it turns out, these coefficients are negligible for $j, k > 6$. For $j, k \leq 6$, they are displayed in the 6×6 matrix below:

$$\begin{pmatrix} 0.8772 & 0.2523 & 0.0684 & 0.0172 & 0.0039 & 0.0000 \\ 0.2523 & 0.1036 & 0.0382 & 0.0131 & 0.0042 & 0.0000 \\ 0.0684 & 0.0382 & 0.0163 & 0.0062 & 0.0022 & 0.0000 \\ 0.0172 & 0.0131 & 0.0062 & 0.0025 & 0.0009 & 0.0000 \\ 0.0039 & 0.0042 & 0.0022 & 0.0009 & 0.0003 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix}\ . \tag{34}$$

Thus, the covariance of the limit of the residual process $\{\widehat{W}_n^2(u) : 0 \leq u \leq 1\}$ of (7) behaves in the form

$$\gamma^2(u,v) = [\min(u,v) - uv] - \frac{2}{\pi^2}\sum_{j=1}^{M}\sum_{k=1}^{M}(\mathbf{a}_2'\mathbf{P}_{jk}\mathbf{a}_2)\frac{\sin(j\pi u)}{j}\frac{\sin(k\pi v)}{k}\ , \tag{35}$$

where $M = 6$. Figure 1 displays all the functions that appear in (35). As seen there, $\gamma^2(u,v)$ is much smaller than the covariance function of the Brownian bridge, because of the substantial correction provided by the second summand at the right-hand side of (35).

The asymptotic variances of the statistics $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, where $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$ is the adjusted residual trace, are equal to $1 -$ the kth diagonal element of the matrix in (34), $k = 1, \ldots, 6$. These appear in table 1. According to table 1, only the first asymptotic variance is really below 1. This corresponds to the leading diagonal element at the upper left corner of the matrix of (34), that takes the value 0.8772. It can be checked numerically that

$$\mathbf{P}_{11} = \begin{pmatrix} 0.8233 & 0.0000 & -0.0590 & 0.0000 \\ 0.0000 & 0.8233 & 0.0000 & -0.0590 \\ -0.0590 & 0.0000 & 0.9311 & 0.0000 \\ 0.0000 & -0.0590 & 0.0000 & 0.9311 \end{pmatrix}\ ; \tag{36}$$

lag	var
1	0.1228
2	0.8964
3	0.9837
4	0.9975
5	0.9997
6	1.0000

Table 1: Asymptotic variances of the leading statistics $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots, 6$, in the bivariate *VAR(1)* model (28)

and

$$\mathbf{P}_{22} = \begin{pmatrix} 0.1107 & 0.0000 & 0.0839 & 0.0000 \\ 0.0000 & 0.1107 & 0.0000 & 0.0839 \\ 0.0839 & 0.0000 & 0.1235 & 0.0000 \\ 0.0000 & 0.0839 & 0.0000 & 0.1235 \end{pmatrix} . \tag{37}$$

Hence, $\mathbf{a}_2'\mathbf{P}_{11}\mathbf{a}_2 = 0.8772$, and $\mathbf{a}_2'\mathbf{P}_{22}\mathbf{a}_2 = 0.1036$. The structure of both $\mathbf{P}_{11}$ and $\mathbf{P}_{22}$ can be explained by taking into account that

$$\mathbf{I}^{-1}[\mathrm{vec}(\mathbf{\Phi}_1)] \cong [\sum_{J=0}^{5} \mathbf{\Phi}_1^J \mathbf{\Sigma} (\mathbf{\Phi}_1^J)']^{-1} \otimes \mathbf{\Sigma} \; . \tag{38}$$

Also, $\mathbf{\Xi}_j' = \mathbf{\Sigma}^{1/2}(\mathbf{\Phi}_1^{j-1})' \otimes \mathbf{\Sigma}^{-1/2}$, $j \geq 1$. Hence, using (38)

$$\mathbf{P}_{jj} = \mathbf{\Xi}_j' \mathbf{I}^{-1}[\mathrm{vec}(\mathbf{\Phi}_1)] \mathbf{\Xi}_j = \mathbf{A}_j \otimes \mathbf{I}_2 \; , \tag{39}$$

where $\mathbf{A}_j \cong \mathbf{\Sigma}^{1/2}(\mathbf{\Phi}_1^{j-1})'[\sum_{J=0}^{5} \mathbf{\Phi}_1^J \mathbf{\Sigma} (\mathbf{\Phi}_1^J)']^{-1} \mathbf{\Phi}_1^{j-1} \mathbf{\Sigma}^{1/2}$, $j \geq 1$. Expression (39) explains the pattern of both the matrices $\mathbf{P}_{11}$ and $\mathbf{P}_{22}$ in (36) and (37), respectively. In particular, it can be checked that

$$\mathbf{A}_1 = \begin{pmatrix} 0.8233 & -0.0590 \\ -0.0590 & 0.9311 \end{pmatrix} \cong \mathbf{I}_2 \quad , \quad \mathbf{A}_2 = \begin{pmatrix} 0.1107 & 0.0839 \\ 0.0839 & 0.1235 \end{pmatrix} .$$

The considerations above can be extended for simulating *VAR(1)* models in dimension $m > 2$. For example, suppose starting eigenvalues $\delta_1 = 0.8500 + 0.1936\,i$; $\delta_2 = 0.8500 - 0.1936\,i$; and $\delta_3 = 0.4359$. The corresponding eigenvectors are taken as

$\boldsymbol{\gamma}_1 = (2.7071, 0.1768 + 0.6847\,i, 1.0000)'$; $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (2.7071, 0.1768 - 0.6847\,i, 1.0000)'$; and $\boldsymbol{\gamma}_3 = (1.0000, 1.0000, 1.0000)'$. Putting $\mathbf{C} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3)$; $\mathbf{D} = \mathrm{diag}(\delta_1, \delta_2, \delta_3)$; and rescaling the expression $\mathbf{CDC}^{-1}$ by dividing by twice its Euclidean norm, leads to the 3×3 matrix

$$\boldsymbol{\Phi}_1 = \begin{pmatrix} 0.2673 & 0.1400 & -0.3275 \\ 0.0346 & 0.1646 & -0.1194 \\ 0.0693 & 0.0517 & -0.0413 \end{pmatrix} . \tag{40}$$

This can be used together with the covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1.0 & 0.3 & 0.3 \\ 0.3 & 1.0 & 0.3 \\ 0.3 & 0.3 & 1.0 \end{pmatrix} , \tag{41}$$

to form a trivariate *VAR(1)* model $\mathbf{X}_t = \boldsymbol{\Phi}_1 \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t$, similar to (28).

Analogue comments to the ones given before for the structure of the parameter space of (28) apply. For instance, the asymptotic variances of the rescaled adjusted residual traces $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{3}$, $k = 1, \ldots, 5$, are displayed in the table below:

lag	var
1	0.0495
2	0.9533
3	0.9973
4	0.9999
5	1.0000

Table 2: Asymptotic variances of the leading statistics $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{3}$, $k = 1, \ldots, 5$, in the trivariate *VAR(1)* model (40)−(41)

4.1.2 Higher order vector autoregressive models

The construction of autoregressive *VAR(p)* models when $p > 1$ requires a criterion for relating the $m \times m$ matrices $\boldsymbol{\Phi}_i$, $i = 1, \ldots, p$, to a collection of prespecified roots, so that they are the solutions of the determinantal equation $|\boldsymbol{\Phi}(z)| = 0$, where $\boldsymbol{\Phi}(z) = \mathbf{I}_m - \boldsymbol{\Phi}_1 z - \cdots - \boldsymbol{\Phi}_p z^p$. The way to proceed is as follows:

(a) For each $j = 1, \ldots, m$, select roots $\varsigma_{j,i}$ with $|\varsigma_{j,i}| > 1$, $i = 1, \ldots, p$.

(b) For each $j = 1, \ldots, m$, form the polynomial of degree p:

$$p_j(z) = 1 - d_{j,1}z - d_{j,2}z^2 - \cdots - d_{j,p}z^p \ , \tag{42}$$

so that its roots are $\varsigma_{j,i}$, $i = 1, \ldots, p$.

(c) Construct the $m \times m$ diagonal matrices

$$\mathbf{D}_i = \text{diag}(d_{1,i}, d_{2,i}, \ldots, d_{m,i}) \ , \quad i = 1, \ldots, p \ . \tag{43}$$

Recall that $\mathbf{D}_i$ is associated to the coefficients of the power z^i in the polynomials $p_j(z)$ of (42), $j = 1, \ldots, m$.

(d) Consider an invertible matrix $\mathbf{A}$ of $m \times m$, and define

$$\mathbf{\Phi}_i = \mathbf{A}\mathbf{D}_i\mathbf{A}^{-1} \ , \quad i = 1, \ldots, p \ . \tag{44}$$

Under the construction (42)−(43)−(44), it follows that

$$|\mathbf{\Phi}(z)| = |\mathbf{I}_m - \mathbf{\Phi}_1 z - \cdots - \mathbf{\Phi}_p z^p| = |\mathbf{I}_m - \mathbf{D}_1 z - \cdots - \mathbf{D}_p z^p| = \prod_{j=1}^{m} p_j(z) \ . \tag{45}$$

By (45), the mp roots of the determinantal equation $|\mathbf{\Phi}(z)| = 0$ are $\varsigma_{j,i}$, $j = 1, \ldots, m$; $i = 1, \ldots, p$. These correspond to those of the polynomials $p_j(z)$ of (42).

j	$\varsigma_{j,1}$	$\lvert\varsigma_{j,1}\rvert$	$\varsigma_{j,2}$	$\lvert\varsigma_{j,2}\rvert$
1	$4.8989 + 4.8989\,i$	6.9281	$4.8989 - 4.8989\,i$	6.9281
2	$7.4282 + 0.0000\,i$	7.4282	$8.9138 + 0.0000\,i$	8.9138
3	$7.9282 + 0.0000\,i$	7.9282	$9.5138 + 0.0000\,i$	9.5138

Table 3: Roots of the determinantal equation $|\mathbf{\Phi}(z)| = 0$ of the trivariate *VAR(2)* model (46)−(41)

As an application of the above algorithm, consider the construction of a trivariate *VAR(2)* model with roots as given in table 3. The coefficients $\{d_{j,i}\}$ of the two degree polynomials $p_j(z) = 1 - d_{j,1}z - d_{j,2}z^2$ of (42) are obtained from the identities:

$$\begin{aligned} d_{j,1} &= \frac{1}{\varsigma_{j,1}} + \frac{1}{\varsigma_{j,2}} \ ; \\ d_{j,2} &= -\frac{1}{\varsigma_{j,1}\varsigma_{j,2}} \ , \end{aligned}$$

$j = 1, 2, 3$. The invertible matrix $\mathbf{A}$ of step **(d)** above is selected in the form

$$\mathbf{A} = \begin{pmatrix} 1.2 & 0.4 & 0.3 \\ 0.3 & 1.0 & 0.3 \\ 0.3 & 0.3 & 1.0 \end{pmatrix} .$$

This leads to the 3×3 matrices:

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.1985 & 0.0180 & 0.0044 \\ -0.0113 & 0.2522 & -0.0029 \\ -0.0089 & 0.0082 & 0.2315 \end{pmatrix} \quad , \quad \mathbf{\Phi}_2 = \begin{pmatrix} -0.0218 & 0.0021 & 0.0019 \\ -0.0018 & -0.0147 & 0.0010 \\ -0.0021 & 0.0001 & -0.0127 \end{pmatrix} . \tag{46}$$

The covariance matrix $\mathbf{\Sigma}$ is taken as in (41).

An important difference with the *VAR(1)* case appears. The leading coefficients $\mathbf{a}_3' \mathbf{P}_{jk} \mathbf{a}_3$ of the second summand of the covariance function are displayed in the 5×5 matrix below:

$$\begin{pmatrix} 0.9997 & 0.0036 & -0.0156 & -0.0036 & 0.0000 \\ 0.0036 & 0.9492 & 0.2156 & 0.0337 & 0.0000 \\ -0.0156 & 0.2156 & 0.0496 & 0.0014 & 0.0000 \\ -0.0036 & 0.0337 & 0.0014 & -0.0002 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix} . \tag{47}$$

It can also be checked numerically that

$$\mathbf{P}_{11} = \begin{pmatrix} 0.9996 & 0.0000 & 0.0000 \\ 0.0000 & 0.9998 & 0.0000 \\ 0.0000 & 0.0000 & 0.9998 \end{pmatrix} \otimes \mathbf{I}_3 \cong \mathbf{I}_9 \ , \tag{48}$$

and

$$\mathbf{P}_{22} = \begin{pmatrix} 0.9591 & -0.0018 & 0.0000 \\ -0.0018 & 0.9392 & -0.0003 \\ 0.0000 & -0.0003 & 0.9471 \end{pmatrix} \otimes \mathbf{I}_3 \cong \mathbf{I}_9 \ . \tag{49}$$

Accordingly, considering the unit vector $\mathbf{a}_3 = \text{vec}(\mathbf{I}_3)/\sqrt{3} = (1, 0, 0, 0, 1, 0, 0, 0, 1)'\sqrt{3}$, it is obtained that both $\mathbf{a}_3' \mathbf{P}_{11} \mathbf{a}_3$ and $\mathbf{a}_3' \mathbf{P}_{22} \mathbf{a}_3$ are close to 1. As seen in figure 2, the correction of the covariance function of the Brownian bridge for $M = 5$ is much stronger than the one observed in figure 1.

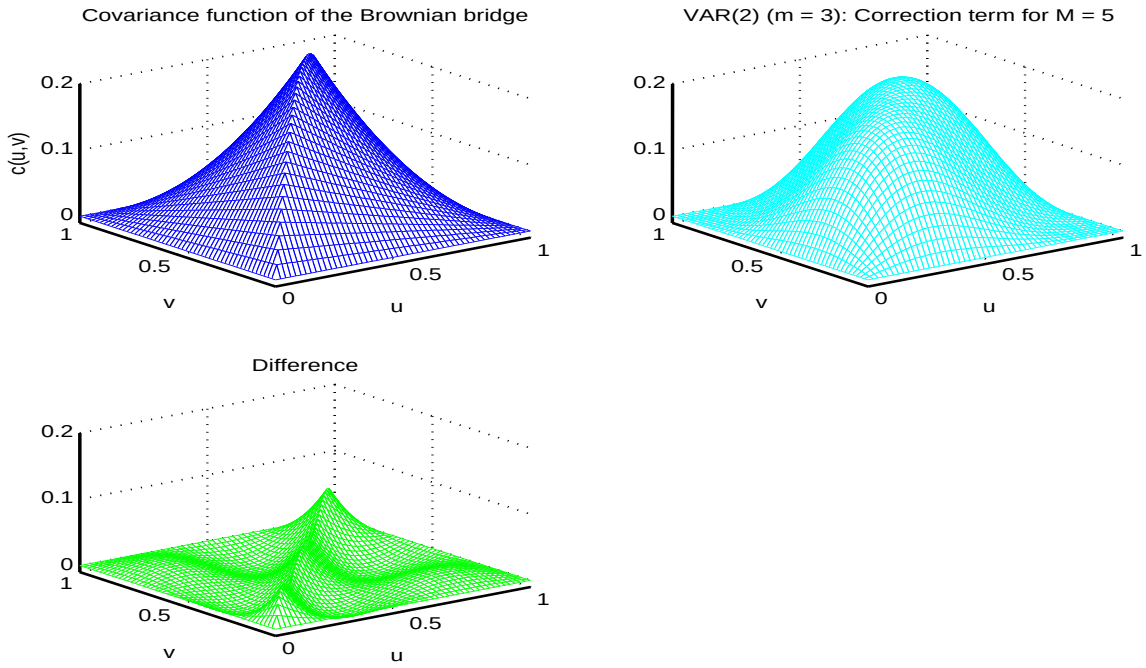

FIG. 2: Covariance function for the limit of the residual process of (7) under the trivariate *VAR(2)* model (46)−(41)

j	$\varsigma_{j,1}$	$\lvert\varsigma_{j,1}\rvert$	$\varsigma_{j,2}$	$\lvert\varsigma_{j,2}\rvert$	$\varsigma_{j,3}$	$\lvert\varsigma_{j,3}\rvert$
1	$4.0000 + 4.0000\,i$	5.6568	$4.0000 - 4.0000\,i$	5.6568	$6.2225 + 0.0000\,i$	6.2225
2	$7.3882 + 0.0000\,i$	7.3882	$8.6196 + 0.0000\,i$	8.6196	$9.8510 + 0.0000\,i$	9.8510

Table 4: Roots of the determinantal equation $|\mathbf{\Phi}(z)| = 0$ of the bivariate *VAR(3)* model (50)−(30)

Another example appears in the construction of a bivariate *VAR(3)* model associated to the roots in table 4. The coefficients $\{d_{j,i}\}$ of the three degree polynomials $p_j(z) = 1 - d_{j,1}z - d_{j,2}z^2 - d_{j,3}z^3$ of (42) are obtained now from the identities:

$$\begin{aligned} d_{j,1} &= \frac{1}{\varsigma_{j,1}} + \frac{1}{\varsigma_{j,2}} + \frac{1}{\varsigma_{j,3}} \ ; \\ d_{j,2} &= -\left(\frac{1}{\varsigma_{j,1}\varsigma_{j,2}} + \frac{1}{\varsigma_{j,1}\varsigma_{j,3}} + \frac{1}{\varsigma_{j,2}\varsigma_{j,3}}\right) ; \\ d_{j,3} &= \frac{1}{\varsigma_{j,1}\varsigma_{j,2}\varsigma_{j,3}} \ , \end{aligned}$$

$j = 1, 2$. The invertible matrix $\mathbf{A}$ of step **(d)** above is selected in the form

$$\mathbf{A} = \begin{pmatrix} 1.2 & 0.4 \\ 0.3 & 1.0 \end{pmatrix} \ .$$

This produces the 2×2 matrices

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.4171 & -0.0257 \\ 0.0161 & 0.3465 \end{pmatrix} \quad ; \quad \mathbf{\Phi}_2 = \begin{pmatrix} -0.0748 & 0.0134 \\ -0.0084 & -0.0379 \end{pmatrix} ;$$

$$\mathbf{\Phi}_3 = \begin{pmatrix} 0.0054 & -0.0015 \\ 0.0010 & 0.0012 \end{pmatrix} . \tag{50}$$

The covariance matrix $\mathbf{\Sigma}$ is as in (30). It can be checked that:

$$\begin{aligned} \mathbf{P}_{11} &= \begin{pmatrix} 0.9972 & 0.0009 \\ 0.0009 & 0.9939 \end{pmatrix} \otimes \mathbf{I}_2 ; \\ \mathbf{P}_{22} &= \begin{pmatrix} 0.9935 & 0.0077 \\ 0.0077 & 0.9808 \end{pmatrix} \otimes \mathbf{I}_2 ; \\ \mathbf{P}_{33} &= \begin{pmatrix} 0.8333 & -0.0051 \\ -0.0051 & 0.8821 \end{pmatrix} \otimes \mathbf{I}_2 , \end{aligned}$$

so that $\mathbf{P}_{jj} \cong \mathbf{I}_4$, $j = 1, 2, 3$. Thus, considering the vector $\mathbf{a}_2 = \text{vec}(\mathbf{I}_2)/\sqrt{2} = (1, 0, 0, 1)'/\sqrt{2}$, for the bivariate *VAR(3)* model (50)−(30) there are now three leading coefficients $\mathbf{a}_2'\mathbf{P}_{jj}\mathbf{a}_2$ close to 1, $j = 1, 2, 3$. After some algebra it can be checked that

$$\mathbf{P}_{jj} = \mathbf{\Xi}_j' \mathbf{I}^{-1}[\text{vec}(\mathbf{\Phi})]\mathbf{\Xi}_j = \mathbf{A}_j \otimes \mathbf{I}_m , \tag{51}$$

where $\mathbf{A}_j \cong \mathbf{I}_m$, $j = 1, \ldots, p$.

An additional $m = 2$ *VAR(2)* model that will be used later is given by the roots in table 5. This produces the 2×2 matrices:

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.2447 & 0.0212 \\ -0.0133 & 0.3031 \end{pmatrix} \quad , \quad \mathbf{\Phi}_2 = \begin{pmatrix} -0.0323 & 0.0041 \\ -0.0026 & -0.0210 \end{pmatrix} . \tag{52}$$

The covariance matrix $\mathbf{\Sigma}$ is taken as in (30).

4.1.3 *VMA(q)* models

The generation of *VMA(1)* models can be treated as the *VAR(1)* case. For instance, taking eigenvalues $\delta_j = 0.0901 \pm 0.0433i$, $j = 1, 2$, with $|\delta_1| = |\delta_2| = 0.0999 < 1$; and

j	$\varsigma_{j,1}$	$\lvert\varsigma_{j,1}\rvert$	$\varsigma_{j,2}$	$\lvert\varsigma_{j,2}\rvert$
1	$4.0000 + 4.0000\,i$	5.6568	$4.0000 - 4.0000\,i$	5.6568
2	$6.1569 + 0.0000\,i$	6.1569	$7.3882 + 0.0000\,i$	7.3882

Table 5: Roots of the determinantal equation $|\mathbf{\Phi}(z)| = 0$ of the bivariate *VAR(2)* model (52)−(30)

eigenvectors $\boldsymbol{\gamma}_1 = (2.7071, 0.2768 + 0.3847i)'$, $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (2.7071, 0.2768 - 0.3847i)'$, the identity $\mathbf{\Theta}_1 = \mathbf{C}\mathbf{D}\mathbf{C}^{-1}$, where $\mathbf{C} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ and $\mathbf{D} = \mathrm{diag}(\delta_1, \delta_2)$, leads to:

$$\mathbf{\Theta}_1 = \begin{pmatrix} 0.0589 & 0.3047 \\ -0.0093 & 0.1212 \end{pmatrix} . \tag{53}$$

The covariance matrix of the errors is selected as

$$\mathbf{\Sigma} = \begin{pmatrix} 1.0 & 0.2 \\ 0.2 & 1.0 \end{pmatrix} . \tag{54}$$

Expressions (53) and (54) can be used to form a *VMA(1)* process of the form $\mathbf{X}_t = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1\boldsymbol{\varepsilon}_{t-1}$. The roots $\varsigma_{j,1}$, $j = 1, 2$, of the determinantal equation $|\mathbf{\Theta}(z)| = |\mathbf{I}_m + \mathbf{\Theta}_1 z| = 0$ are related to the eigenvalues in the form $\varsigma_{j,1} = -\delta_j$, $j = 1, 2$.

For the bivariate *VMA(1)* model (53)−(54), the asymptotic variances of the adjusted residual traces $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots, 4$ appear in table 6.

lag	var
1	0.0498
2	0.9516
3	0.9986
4	1.0000

Table 6: Asymptotic variances of the leading statistics $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots, 5$, in the bivariate *VMA(1)* model (53)−(54)

In this case, only the coefficient $\mathbf{a}_2'\mathbf{P}_1\mathbf{a}_2$ is close enough to 1. It can be also checked numerically that

$$\mathbf{P}_{11} = \mathbf{\Xi}_1'\mathbf{I}^{-1}[\mathrm{vec}(\mathbf{\Theta}_1)]\mathbf{\Xi}_1 = \mathbf{I}_2 \otimes \begin{pmatrix} 0.9925 & -0.0243 \\ -0.0243 & 0.9080 \end{pmatrix} =$$

$$= \begin{pmatrix} 0.9925 & -0.0243 & 0.0000 & 0.0000 \\ -0.0243 & 0.9080 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.9925 & -0.0243 \\ 0.0000 & 0.0000 & -0.0243 & 0.9080 \end{pmatrix} \cong \mathbf{I}_4 \ .$$

j	$\varsigma_{j,1}$	$\lvert\varsigma_{j,1}\rvert$	$\varsigma_{j,2}$	$\lvert\varsigma_{j,2}\rvert$
1	$2.0000 + 2.0000\,i$	2.8284	$2.0000 - 2.0000\,i$	2.8284
2	$3.3284 + 0.0000\,i$	3.3284	$3.9941 + 0.0000\,i$	3.9941

Table 7: Roots of the determinantal equation $|\mathbf{\Theta}(z)| = 0$ of the bivariate *VMA(2)* model (55)−(54)

The construction of *VMA(q)* processes for $q > 1$ proceeds by adapting conveniently the method for *VAR(p)* models when $p > 1$. For instance, suppose that it is desired to form a bivariate *VMA(2)* model with roots as given in table 7. The coefficients $\{d_{j,i}\}$ of the two degree polynomials $p_j(z) = 1 + d_{j,1}z + d_{j,2}z^2$ of (42) are determined now from the identities:

$$\begin{aligned} d_{j,1} &= -\frac{1}{\varsigma_{j,1}} - \frac{1}{\varsigma_{j,2}} \ ; \\ d_{j,2} &= \frac{1}{\varsigma_{j,1}\varsigma_{j,2}} \ , \end{aligned}$$

$j = 1, 2$. The invertible matrix $\mathbf{A}$ of step **(d)** is

$$\mathbf{A} = \begin{pmatrix} 1.2 & 0.4 \\ 0.2 & 1.0 \end{pmatrix} \ .$$

This leads to the 2×2 matrices:

$$\mathbf{\Theta}_1 = \begin{pmatrix} -0.4964 & -0.0218 \\ 0.0091 & -0.5544 \end{pmatrix} \quad , \quad \mathbf{\Theta}_2 = \begin{pmatrix} 0.1286 & -0.0213 \\ 0.0089 & 0.0717 \end{pmatrix} \ . \tag{55}$$

The covariance matrix of the errors $\mathbf{\Sigma}$ is as in expression (54). For this *VMA(2)* model $\mathbf{X}_t = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1\boldsymbol{\varepsilon}_{t-1} + \mathbf{\Theta}_2\boldsymbol{\varepsilon}_{t-2}$ defined by (55)−(54), it is found that

$$\begin{aligned} \mathbf{P}_{11} &= \mathbf{I}_2 \otimes \begin{pmatrix} 0.9841 & 0.0018 \\ 0.0018 & 0.9841 \end{pmatrix} \ , \\ \mathbf{P}_{22} &= \mathbf{I}_2 \otimes \begin{pmatrix} 0.7677 & -0.0073 \\ -0.0073 & 0.7116 \end{pmatrix} \ . \end{aligned}$$

The asymptotic variances of the adjusted residual traces $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots, 7$, are displayed in table 8.

lag	var
1	0.0108
2	0.2369
3	0.7840
4	0.9882
5	0.9996
6	1.0000
7	1.0000

Table 8: Asymptotic variances of the leading statistics $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots, 7$, in the bivariate *VMA(2)* model (55)−(54)

4.1.4 *VARMA(p,q)* models

The construction of *VARMA(p,q)* processes can be done by combining the rules presented earlier for the autoregressive and moving average parts of the model, respectively. For example, a bivariate *VARMA(1,1)* process of the form

$$\mathbf{X}_t - \mathbf{\Phi}_1 \mathbf{X}_{t-1} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1 \boldsymbol{\varepsilon}_{t-1} \tag{56}$$

can be obtained using the matrices $\mathbf{\Phi}_1$ and $\mathbf{\Theta}_1$ of expressions (29) and (53) respectively, and the covariance matrix $\mathbf{\Sigma}$ of (54). In this case, it is found that

$$\mathbf{P}_{11} = \begin{pmatrix} 0.9986 & -0.0045 & -0.0005 & -0.0016 \\ -0.0045 & 0.9829 & -0.0016 & -0.0062 \\ -0.0005 & -0.0016 & 0.9995 & -0.0017 \\ -0.0016 & -0.0062 & -0.0017 & 0.9934 \end{pmatrix},$$

and

$$\mathbf{P}_{22} = \begin{pmatrix} 0.8680 & 0.0549 & -0.0454 & 0.0654 \\ 0.0549 & 0.7906 & -0.0070 & -0.0397 \\ -0.0454 & -0.0070 & 0.9683 & 0.0451 \\ 0.0654 & -0.0397 & 0.0451 & 0.8948 \end{pmatrix}.$$

Both matrices above are reasonably close to the identity $\mathbf{I}_4$. The asymptotic variances of the adjusted residual traces $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots, 7$, are displayed in table 9 below. According to this numerical information, only the first two scaled adjusted residual traces have a variance markedly smaller than 1.

lag	var
1	0.0056
2	0.0532
3	0.9520
4	0.9906
5	0.9988
6	0.9998
7	1.0000

Table 9: Asymptotic variances of the leading statistics $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots, 7$, in the bivariate *VARMA(1,1)* model (56)−(54)

By considering for the determinantal equation $|\mathbf{\Phi}(z)| = |\mathbf{I}_m - \mathbf{\Phi}_1 z - \mathbf{\Phi}_2 z^2| = 0$ the roots $\varsigma_{1,1} = 5.9999 + 5.9999i$; $\varsigma_{1,2} = 5.9999 - 5.9999i$, so that $|\varsigma_{1,1}| = |\varsigma_{1,2}| = 8.4851$; $\varsigma_{2,1} = 8.9853 + 0.0000i$; and $\varsigma_{2,2} = 10.7823 + 0.0000i$; and using the invertible matrix

$$\mathbf{A} = \begin{pmatrix} 1.2 & 0.4 \\ 0.2 & 1.0 \end{pmatrix} ,$$

it is obtained

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.1640 & 0.0160 \\ -0.0067 & 0.2067 \end{pmatrix} \quad , \quad \mathbf{\Phi}_2 = \begin{pmatrix} -0.0141 & 0.0015 \\ -0.0006 & -0.0101 \end{pmatrix} . \tag{57}$$

Taking in turn roots $\varsigma_{1,1} = 4.0000 + 4.0000i$; $\varsigma_{1,2} = 4.0000 - 4.0000i$, with $|\varsigma_{1,1}| = |\varsigma_{1,2}| = 5.6568$; $\varsigma_{2,1} = 6.1569 + 0.0000i$; and $\varsigma_{2,2} = 7.3882 + 0.0000i$ for the equation $|\mathbf{\Theta}(z)| = |\mathbf{I}_m + \mathbf{\Theta}_1 z + \mathbf{\Theta}_2 z^2| = 0$, and using the same $\mathbf{A}$ as above it is found that

$$\mathbf{\Theta}_1 = \begin{pmatrix} -0.2466 & -0.0205 \\ 0.0085 & -0.3012 \end{pmatrix} \quad , \quad \mathbf{\Theta}_2 = \begin{pmatrix} 0.0319 & -0.0040 \\ 0.0017 & 0.0213 \end{pmatrix} . \tag{58}$$

The 2×2 matrices of (57) and (58), combined with the error covariance matrix $\mathbf{\Sigma}$ of (54), lead finally to a *VARMA(2,2)* process of the form

$$\mathbf{X}_t - \mathbf{\Phi}_1 \mathbf{X}_{t-1} - \mathbf{\Phi}_2 \mathbf{X}_{t-2} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1 \boldsymbol{\varepsilon}_{t-1} + \mathbf{\Theta}_2 \boldsymbol{\varepsilon}_{t-1} \ . \tag{59}$$

4.2 Behavior of the adjusted traces

This section studies and compares the properties of the different versions of the adjusted traces: the residual $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$; the modified $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$; and the model error version $\mathrm{tr}(\mathbf{R}_k)/\sqrt{m}$.

Notice that $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m} \overset{D}{\cong} N(0, 1 - \mathbf{a}_m' \mathbf{P}_{kk} \mathbf{a}_m)$ where $\mathbf{a}_m = \mathrm{vec}(\mathbf{I}_m)/\sqrt{m}$ is a unit $m^2 \times 1$ vector, and $\mathbf{P}_{kk} = \mathbf{\Xi}_k' \mathbf{I}^{-1}(\mathbf{\Lambda}) \mathbf{\Xi}_k$ is the $m^2 \times m^2$ matrix, $k = 1, \ldots, M$. We suggest a plot of the adjusted residual traces $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ with the residual version of bands

$$\pm z_{\alpha/2}\, n^{-1/2} \sqrt{1 - \mathbf{a}_m' \mathbf{\Xi}_k' (\mathbf{Z}_M' \mathcal{W}^{-1} \mathbf{Z}_M)^{-1} \mathbf{\Xi}_k \mathbf{a}_m} \ , \quad 1 \leq k \leq M \ , \tag{60}$$

as a possible diagnostic check in *VARMA(p,q)* processes, where $z_{\alpha/2}$ is a suitable quantile of a $N(0,1)$ distribution. The bands of (60),

$$\pm 1.96\, n^{-1/2} \left(1 - \mathbf{a}_m' \mathbf{P}_{kk} \mathbf{a}_m\right)^{1/2} \ , \quad 1 \leq k \leq M \ ,$$

corresponding to a sample size of $n = 250$ and a nominal level $\alpha = .05$, are plotted in figure 3 for seven bivariate time series models:

(a) The *VAR(1)* of expression (28);

(b) The *VAR(2)* of expressions (52)−(30);

(c) The *VAR(3)* of expressions (50)−(30);

(d) The *VMA(1)* of expressions (53)−(54);

(e) The *VMA(2)* of expressions (55)−(54);

(f) The *VARMA(1,1)* of expression (56); and

(g) The *VARMA(2,2)* of expression (59).

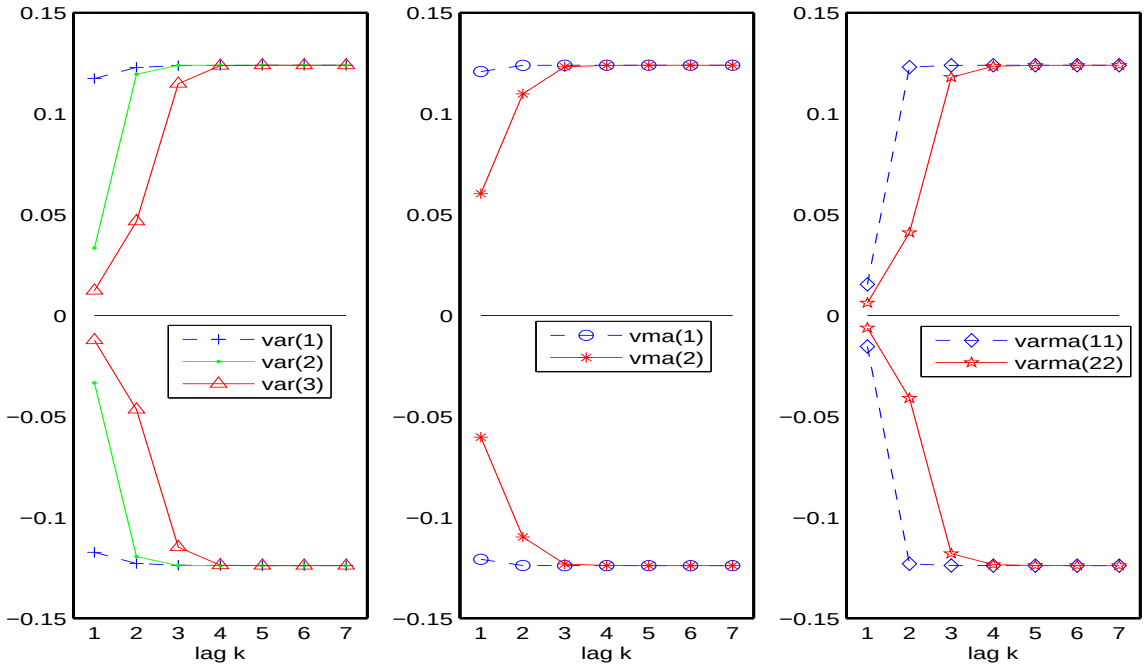

FIG. 3: Bands $\pm 1.96 n^{-1/2}(1 - \mathbf{a}_m' \mathbf{P}_{kk} \mathbf{a}_m)^{1/2}$, $k = 1, \ldots, M$, with $n = 250$ for the seven models

The message of figure 3 is confirmed numerically by the values of the asymptotic variances $1 - \mathbf{a}_m' \mathbf{P}_{kk} \mathbf{a}_m$, $k = 1, \ldots, M$. These appear in table 10 below:

lag	*VAR(1)*	*VAR(2)*	*VAR(3)*	*VMA(1)*	*VMA(2)*	*VARMA(1,1)*	*VARMA(2,2)*
1	0.1228	0.0007	0.0016	0.0498	0.0108	0.0056	0.0001
2	0.8964	0.0727	0.0099	0.9516	0.2369	0.0532	0.0001
3	0.9837	0.9293	0.1413	0.9986	0.7840	0.9520	0.0002
4	0.9975	0.9999	0.8573	1.0000	0.9882	0.9906	0.1258
5	0.9997	1.0000	0.9980	–	0.9996	0.9988	0.9166
6	1.0000	–	0.9999	–	1.0000	0.9998	0.9582
7	–	–	1.0000	–	1.0000	1.0000	0.9985
8	–	–	–	–	–	–	0.9999

Table 10: Asymptotic variances of the leading statistics $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{2}$, $k = 1, \ldots, M$, for the seven models

In order to compare the behavior of the different adjusted traces, $N = 1000$ independent replicas of size $n = 250$ are generated from the bivariate *VAR(1)*, *VMA(1)*, and *VARMA(1,1)* models considered in figure 3 and table 10. The results are presented in

the histograms that appear in figures 4, 5, and 6, respectively. For ease of presentation, an axis label tr1 refers to the rescaled adjusted residual trace $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_1)/\sqrt{m}$; a label ts1 to the modified $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{S}}_{p+q+1})/\sqrt{m}$; and a label te1 to the model error version $\sqrt{n}\,\mathrm{tr}(\mathbf{R}_1)/\sqrt{m}$.

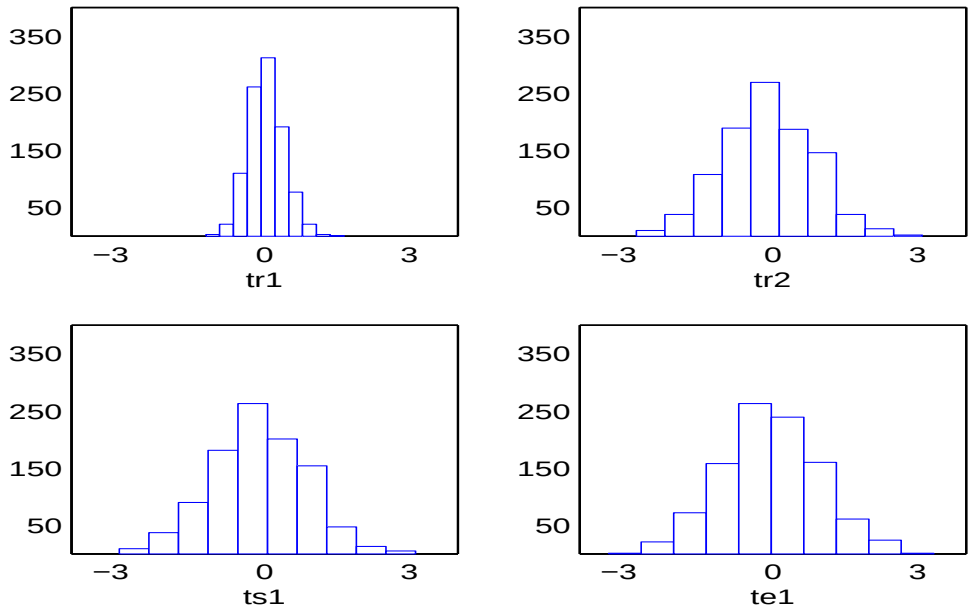

FIG. 4: Histograms of the adjusted traces for $N = 1000$ independent replicas of size $n = 250$ for the bivariate model *VAR(1)* (28)

In general, both $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ and $\sqrt{n}\,\mathrm{tr}(\mathbf{R}_k)/\sqrt{m}$ are close to a $N(0,1)$. On the other hand, for low values of the lag k, the rescaled adjusted residual traces $\sqrt{n}\,\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ behave like a centered normal with a variance smaller than 1. The variance goes to 1 when k increases, and thus the original adjusted residual traces will tend to behave as the modified ones.

The aforementioned analogy holds not only in distribution, but also numerically. This is because for k and n large enough

$$\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m} = \mathbf{a}_m'\mathrm{vec}(\widehat{\mathbf{S}}_k) \cong \mathbf{a}_m'\mathrm{vec}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{C}}_k'\widehat{\mathbf{\Sigma}}^{-1/2}) = \mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}\ .$$

Table 11 displays the values of the statistics $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$ and $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ for the last generated sample of the simulation experiment that produces the histograms in figures 4, 5, and 6. Only the values of the adjusted traces that are different are presented.

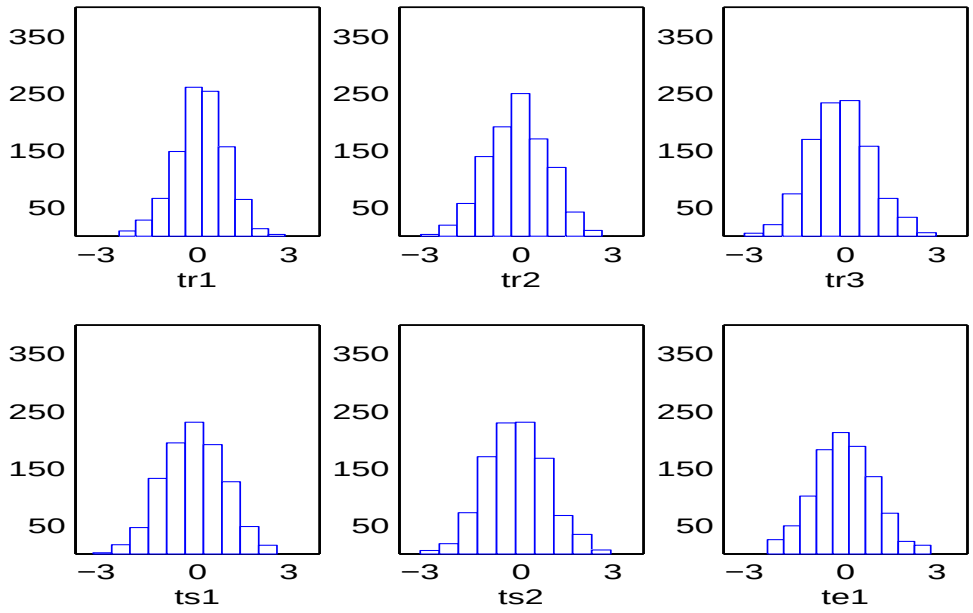

FIG. 5: Histograms of the adjusted traces for $N = 1000$ independent replicas of size $n = 250$ for the bivariate model *VMA(1)* (53)−(54)

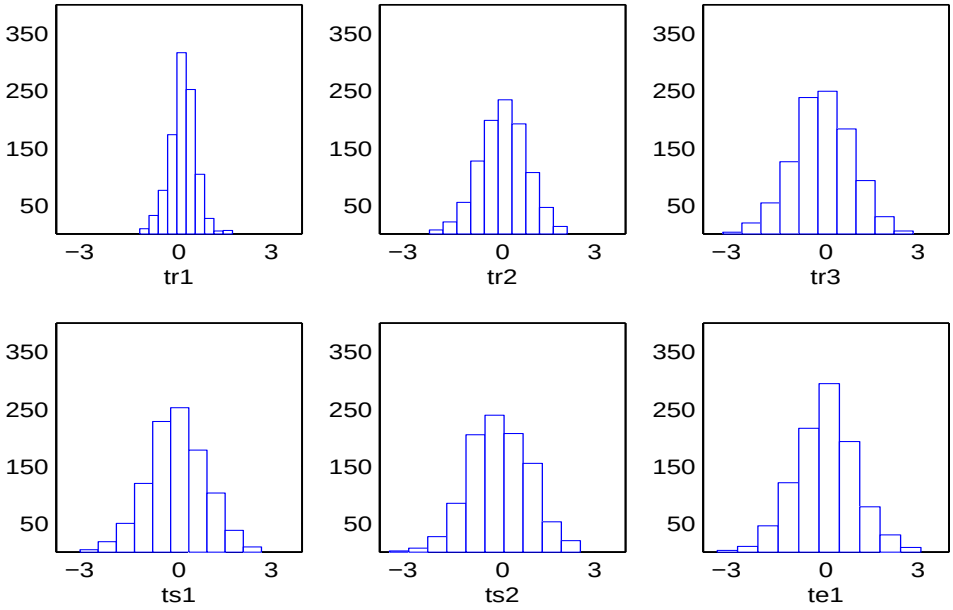

FIG. 6: Histograms of the adjusted traces for $N = 1000$ independent replicas of size $n = 250$ for the bivariate model *VARMA(1,1)* (56)

From table 11, just the first traces differ to some extent. As it may be expected, the differences, that are always moderate in size, increase with the complexity of the model.

	VAR(1)		*VMA(1)*		*VARMA(1,1)*	
lag	$\frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}$	$\frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}$	$\frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}$	$\frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}$	$\frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}$	$\frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}$
1	0.0037	–	−0.0370	–	−0.0002	–
2	0.0908	0.0754	−0.0345	−0.0479	−0.0033	–
3	−0.0569	−0.0630	0.1574	0.1565	−0.1090	−0.1047
4	−0.0524	−0.0531	−0.1706	−0.1700	0.0546	0.0505
5	−0.0896	−0.0896	−0.0837	−0.0837	0.1633	0.1652
6	···	···	···	···	−0.1800	−0.1803
7	···	···	···	···	−0.0009	−0.0007
8	···	···	···	···	−0.0319	−0.0320

Table 11: Original and modified adjusted residual traces for the last generated sample of size $n = 250$ in the experiment underlying the histograms appearing in figures 4, 5, and 6

4.3 Comparisons between goodness-of-fit processes

Goodness-of-fit functionals considered are the Kolmogorov-Smirnov criterion,

$$\sup_{0\leq u\leq 1} |\widehat{Z}_n^m(u)| \; ; \tag{61}$$

and the Cramér-von Mises statistic,

$$CVM = \int_0^1 [\widehat{Z}_n^m(u)]^2 du = \frac{n}{m\,\pi^2} \sum_{k=p+q+1}^{n-(P+1)} \frac{[\mathrm{tr}(\widehat{\mathbf{S}}_k)]^2}{K^2} \; . \tag{62}$$

In practice, criterion (61) is approximated by

$$KS = \sup_{1\leq j\leq n} |\widehat{Z}_n^m(j/n)| \; . \tag{63}$$

Using the tightness condition for the distributions of $\{\widehat{Z}_n^m(u) : 0 \leq u \leq 1\}$, the asymptotic distribution of KS is that of (61), $\sup_{0\leq u\leq 1} |B(u)|$. Another possibility is to approximate the criterion CVM in (62) by the Riemann sum

$$PCVM = \frac{1}{n} \sum_{j=1}^{n} [\widehat{Z}_n^m(j/n)]^2 \; . \tag{64}$$

The limit distribution of both CVM and $PCVM$ is $\int_0^1 [B(u)]^2 du$. For the residual process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ of (7) criterion KS rejects for a nominal level α the adequacy of a given model when

$$KS = \sup_{1 \leq j \leq n} |\widehat{W}_n^m(j/n)| > KS_\alpha \ ,$$

where KS_α is the proper quantile of the distribution of $\sup_{1 \leq u \leq 1} |B(u)|$. Similarly for criteria CVM and $PCVM$.

The statistics KS of (63); CVM of (62); and $PCVM$ of (64) are compared to the adequate critical points of the distributions of $\sup_{0 \leq u \leq 1} |B(u)|$ and $\int_0^1 [B(u)]^2 du$, respectively. For a nominal significance level $0 < \alpha < 1$, the notation for the corresponding $(1 - \alpha) \times 100\%$ quantiles will be KS_α and CVM_α, respectively. Nominal levels $\alpha = .1$, $.05$, and $.01$ will be used. It is obtained that $KS, .1 = 1.2238$; $KS, .05 = 1.3582$; and $KS, .01 = 1.6277$. Also, $CVM, .1 = 0.3473$; $CVM, .05 = 0.4614$; and $CVM, .01 = 0.7435$.

4.3.1 Size

For the bivariate *VAR(1)*, *VMA(1)*, and *VARMA(1,1)* models considered in figure 3; table 10; and histograms 4−5−6, $N = 1000$ independent replicas of size $n = 250$ are generated. Nominal significance levels considered are $\alpha = .1$, and $.05$. Results are presented in table 12 for *VAR(1)*; table 13 for *VMA(1)*; and table 14 for *VARMA(1,1)*. For a given value of α, the information contained in the columns of these tables is as follows:

(a) Empirical proportion of rejections, say $\widehat{p}_N$;

(b) Lower bound of a 95% confidence interval for the true probability of rejection at level α, $\widehat{p}_N - 1.96\sqrt{\widehat{p}_N(1 - \widehat{p}_N)/N}$;

(c) Upper bound version of **(b)**, $\widehat{p}_N + 1.96\sqrt{\widehat{p}_N(1 - \widehat{p}_N)/N}$;

(d) Theoretical quantile of the specific criterion;

(e) Empirical quantile for the $N = 1000$ replicas simulated.

As a conclusion from tables 12−13−14, the empirical size of the criteria based on functionals applied on the error and modified processes is quite close to the nominal. As expected, the behavior of CVM and $PCVM$ is almost identical. Consequently, $PCVM$ will be ignored from now on. However, when considering KS, CVM, and $PCVM$ on the residual process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ of (7), the size of the regions of the form $H[\widehat{W}_n^m(u)] \geq H_\alpha[B(u)]$, is well below α. This indicates that the consequences of an inequality in Anderson (1955) can be very severe in practice.

4.3.2 Power

For making power comparisons between the functionals applied on the residual process $\{\widehat{W}_n^m(u) : 0 \leq u \leq 1\}$ and the modified process $\{\widehat{Z}_n^m(u) : 0 \leq u \leq 1\}$, a bivariate *VARMA(1,1)* model of the form $\mathbf{X}_t - \mathbf{\Phi}_1\mathbf{X}_{t-1} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1(\beta)\boldsymbol{\varepsilon}_{t-1}$ is considered, where $\mathbf{\Theta}_1(\beta) = \beta\mathbf{\Theta}_1$; β is a real parameter in the interval $[0, 1)$;

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.1868 & 0.1787 \\ -0.0122 & 0.2101 \end{pmatrix} ; \tag{65}$$

and

$$\mathbf{\Theta}_1 = \begin{pmatrix} 1.3008 & -0.1945 \\ 1.7858 & 0.5017 \end{pmatrix} . \tag{66}$$

The matrix $\mathbf{\Phi}_1$ of (65) is obtained by taking eigenvalues $\delta_j = 0.1984 \pm 0.0452\, i$, $j = 1, 2$, so that $|\delta_1| = |\delta_2| = 0.2035 < 1$. The associated eigenvectors are $\boldsymbol{\gamma}_1 = (0.9675, 0.0632 + 0.2447\, i)'$; $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (0.9675, 0.0632 - 0.2447\, i)'$. The eigenvalues of the matrix $\mathbf{\Theta}_1$ of (66) are $\delta_j = 0.9013 \pm 0.4333\, i$, $j = 1, 2$, so that $|\delta_1| = |\delta_2| = 1.0000$. The eigenvectors are $\boldsymbol{\gamma}_1 = (0.2125 + 0.2304\, i, 0.9496)'$, and $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (0.2125 - 0.2304\, i, 0.9496)'$. The covariance matrix $\mathbf{\Sigma}$ of the errors is as in (54).

For each value in a grid of values of the parameter $0 \leq \beta < 1$, $N = 1000$ independent data samples of length $n = 200$ are generated from the process $\mathbf{X}_t - \mathbf{\Phi}_1\mathbf{X}_{t-1} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1(\beta)\boldsymbol{\varepsilon}_{t-1}$ defined by the matrices in (65)−(66)−(54). If a *VAR(1)* process is postulated and fitted, the value of $\beta = 0$ corresponds to the null *VAR(1)* model. The values of $0 < \beta < 1$ define an alternative *VARMA(1,1)* model. Thus, the plots of the corresponding empirical proportions of rejections at level $\alpha = .05$ versus β give a graphical display of

the power function of the method. The modulus of the eigenvalues of $\mathbf{\Theta}_1(\beta)$ is β, so that when $\beta \to 1$ the alternative *VARMA(1,1)* process is close to having a unit root. The results are displayed at the left part of figure 7. For moderate values of β, both KS and CVM on the modified process are more powerful than the same functionals on the original process All the power functions tend to unity when $\beta \to 1$. The larger power of KS and CVM for moderate to large values of β in the original residual process can be explained by the fact that the $m \times m$ residual correlation matrices $\widehat{\mathbf{R}}_k$ are based on residual vectors $\widehat{\boldsymbol{\varepsilon}}_t$ that are more sensitive to departures form the null assumption than the modified matrices $\widehat{\mathbf{S}}_k$.

An additional experiment is conducted. A $m = 2$ *VARMA(2,2)* process of the form $\mathbf{X}_t - \mathbf{\Phi}_1\mathbf{X}_{t-1} - \mathbf{\Phi}_2\mathbf{X}_{t-2} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1(\beta)\boldsymbol{\varepsilon}_{t-1} + \mathbf{\Theta}_2(\beta)\boldsymbol{\varepsilon}_{t-2}$ is considered, where $0 \leq \beta < 1$. The 2×2 matrices $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are obtained with the method of section 4.1.2. The roots of the determinantal equation $|\mathbf{\Phi}(z)| = |\mathbf{I}_2 - \mathbf{\Phi}_1 z - \mathbf{\Phi}_2 z^2| = 0$ are given by $\varsigma_{1,1} = 0.7589 + 1.5178\,i$, $\varsigma_{1,2} = 0.7589 - 1.5178\,i$, so that $|\varsigma_{1,i}| = 1.6979$, $i = 1, 2$; and $\varsigma_{2,1} = 2.1971 + 0.0000\,i$, $\varsigma_{2,2} = 2.8562 + 0.0000\,i$. The invertible matrix

$$\mathbf{A} = \begin{pmatrix} 1.8 & 0.7 \\ 0.2 & 1.0 \end{pmatrix}$$

leads to

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.5036 & 0.2112 \\ -0.0335 & 0.8287 \end{pmatrix} \quad , \quad \mathbf{\Phi}_2 = \begin{pmatrix} -0.3631 & 0.1426 \\ -0.0226 & -0.1435 \end{pmatrix} . \tag{67}$$

On the other hand, the matrices $\mathbf{\Theta}_1(\beta)$ and $\mathbf{\Theta}_2(\beta)$ are selected so that $\mathbf{\Theta}_1(0) = \mathbf{\Theta}_2(0) = \mathbf{0}_{2\times 2}$. For $0 < \beta < 1$, the roots of the determinantal equation $|\mathbf{I}_2 + \mathbf{\Theta}_1(\beta)z + \mathbf{\Theta}_2(\beta)z^2| = 0$ are of the form

$$\varsigma_{j,i}(\beta) = q(\beta)\,\varsigma_{j,i} \ , \quad j, i = 1, 2 \ , \tag{68}$$

where $q(\beta) = 2.8469 - 1.8469\,\beta$ for $0 < \beta < 1$. The $\varsigma_{j,i}$ are as specified in table 15 below. The covariance matrix $\mathbf{\Sigma}$ of the errors is taken again as in (54). The invertible matrix of the method is now

$$\mathbf{A} = \begin{pmatrix} 2.2 & 0.4 \\ 0.2 & 1.0 \end{pmatrix} .$$

$\alpha = .10$	**(a)**	**(b)**	**(c)**	**(d)**	**(e)**
	Error process				
KS	0.0650	0.0497	0.0803	1.2238	1.1261
CVM	0.0990	0.0805	0.1175	0.3473	0.3395
PCVM	0.0990	0.0805	0.1175	0.3473	0.3395
	Residual process				
KS	0.0030	-0.0004	0.0064	1.2238	0.8191
CVM	0.0000	0.0000	0.0000	0.3473	0.1369
PCVM	0.0000	0.0000	0.0000	0.3473	0.1369
	Modified process				
KS	0.0860	0.0686	0.1034	1.2238	1.1864
CVM	0.1090	0.0897	0.1283	0.3473	0.3494
PCVM	0.1090	0.0897	0.1283	0.3473	0.3494
$\alpha = .05$	**(a)**	**(b)**	**(c)**	**(d)**	**(e)**
	Error process				
KS	0.0330	0.0219	0.0441	1.3582	1.2549
CVM	0.0490	0.0356	0.0624	0.4614	0.4476
PCVM	0.0490	0.0356	0.0624	0.4614	0.4476
	Residual process				
KS	0.0000	0.0000	0.0000	1.3582	0.9343
CVM	0.0000	0.0000	0.0000	0.4614	0.1641
PCVM	0.0000	0.0000	0.0000	0.4614	0.1641
	Modified process				
KS	0.0390	0.0270	0.0510	1.3582	1.3284
CVM	0.0440	0.0313	0.0567	0.4614	0.4443
PCVM	0.0440	0.0313	0.0567	0.4614	0.4443

Table 12: Empirical sizes for $N = 1000$ independent replicas of size $n = 250$ for the bivariate model *VAR(1)* (28)

$\alpha = .10$	**(a)**	**(b)**	**(c)**	**(d)**	**(e)**
	Error process				
KS	0.0660	0.0506	0.0814	1.2238	1.1440
CVM	0.1090	0.0897	0.1283	0.3473	0.3619
PCVM	0.1090	0.0897	0.1283	0.3473	0.3619
	Residual process				
KS	0.0350	0.0236	0.0464	1.2238	1.0150
CVM	0.0420	0.0296	0.0544	0.3473	0.2364
PCVM	0.0420	0.0296	0.0544	0.3473	0.2364
	Modified process				
KS	0.0840	0.0668	0.1012	1.2238	1.1801
CVM	0.0870	0.0695	0.1045	0.3473	0.3321
PCVM	0.0870	0.0695	0.1045	0.3473	0.3321
$\alpha = .05$	**(a)**	**(b)**	**(c)**	**(d)**	**(e)**
	Error process				
KS	0.0270	0.0170	0.0370	1.3582	1.2738
CVM	0.0520	0.0382	0.0658	0.4614	0.4621
PCVM	0.0520	0.0382	0.0658	0.4614	0.4621
	Residual process				
KS	0.0120	0.0053	0.0187	1.3582	1.1503
CVM	0.0170	0.0090	0.0250	0.4614	0.3162
PCVM	0.0170	0.0090	0.0250	0.4614	0.3162
	Modified process				
KS	0.0430	0.0304	0.0556	1.3582	1.3398
CVM	0.0450	0.0322	0.0578	0.4614	0.4455
PCVM	0.0450	0.0322	0.0578	0.4614	0.4455

Table 13: Empirical sizes for $N = 1000$ independent replicas of size $n = 250$ for the bivariate model *VMA(1)* (53)−(54)

$\alpha = .10$	**(a)**	**(b)**	**(c)**	**(d)**	**(e)**
	Error process				
KS	0.0570	0.0426	0.0714	1.2238	1.1314
CVM	0.0930	0.0750	0.1110	0.3473	0.3331
PCVM	0.0930	0.0750	0.1110	0.3473	0.3331
	Residual process				
KS	0.0030	-0.0004	0.0064	1.2238	0.8145
CVM	0.0020	-0.0008	0.0048	0.3473	0.1235
PCVM	0.0020	-0.0008	0.0048	0.3473	0.1235
	Modified process				
KS	0.0770	0.0605	0.0935	1.2238	1.1613
CVM	0.0780	0.0614	0.0946	0.3473	0.3122
PCVM	0.0780	0.0614	0.0946	0.3473	0.3122
$\alpha = .05$	**(a)**	**(b)**	**(c)**	**(d)**	**(e)**
	Error process				
KS	0.0300	0.0194	0.0406	1.3582	1.2356
CVM	0.0490	0.0356	0.0624	0.4614	0.4419
PCVM	0.0490	0.0356	0.0624	0.4614	0.4419
	Residual process				
KS	0.0000	0.0000	0.0000	1.3582	0.8871
CVM	0.0000	0.0000	0.0000	0.4614	0.1470
PCVM	0.0000	0.0000	0.0000	0.4614	0.1470
	Modified process				
KS	0.0330	0.0219	0.0441	1.3582	1.2903
CVM	0.0420	0.0296	0.0544	0.4614	0.4203
PCVM	0.0420	0.0296	0.0544	0.4614	0.4203

Table 14: Empirical sizes for $N = 1000$ independent replicas of size $n = 250$ for the bivariate model *VARMA(1,1)* (56)

j	$\varsigma_{j,1}$	$\lvert\varsigma_{j,1}\rvert$	$\varsigma_{j,2}$	$\lvert\varsigma_{j,2}\rvert$
1	$0.7071 + 0.7071\,i$	1.0000	$0.7071 - 0.7071\,i$	1.0000
2	$1.1095 + 0.0000\,i$	1.1095	$1.1314 + 0.0000\,i$	1.1314

Table 15: Multiples of the roots of the *VMA* part in the parametric bivariate *VARMA(2,2)* model (67)−(68)−(54)

Notice that $q(\beta)$ decreases towards 1 when $\beta \to 1$. Thus, the VMA part of the bivariate *VARMA(2,2)* model defined by (67)−(68)−(54) approaches also to a unit root situation. If a *VAR(2)* is postulated and fitted, the value of $\beta = 0$ corresponds to the null *VAR(2)* process. The values of $0 < \beta < 1$ form an alternative *VARMA(2,2)* model. As before, for each value of a grid of values of $0 \leq \beta < 1$, $N = 1000$ independent data samples of length $n = 200$ are generated from the *VARMA(2,2)* model (67)−(68)−(54). The associated plot of empirical powers at level $\alpha = .05$ is given at the right of figure 7, in which the functionals based on the modified process are clearly much more powerful.

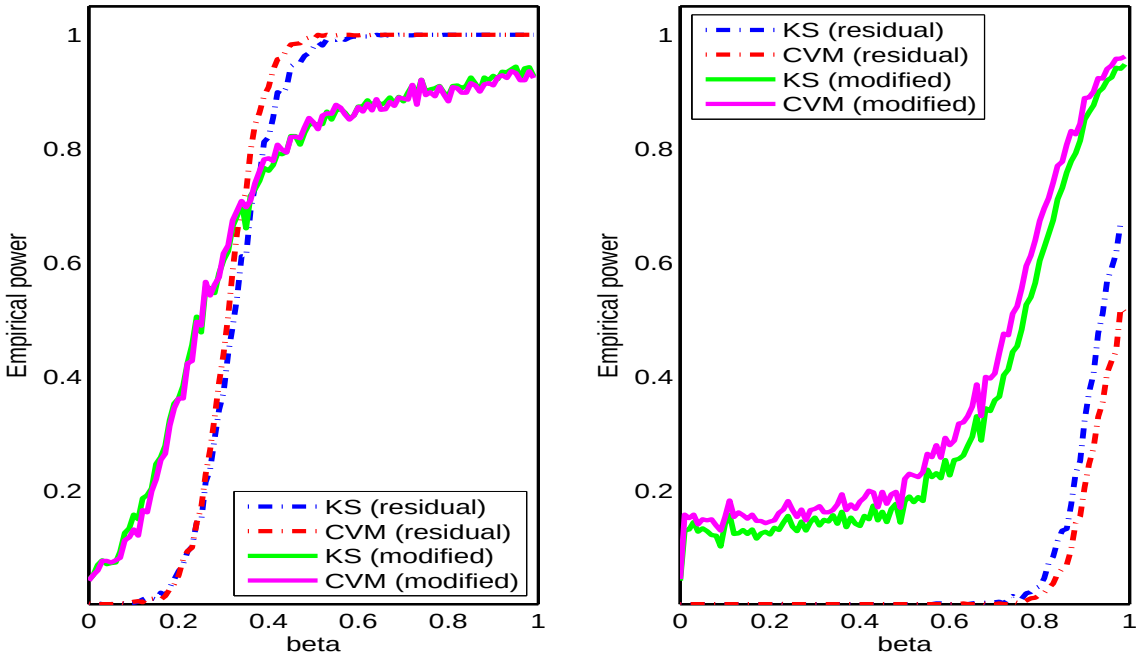

FIG. 7: Comparison of empirical powers of the residual and modified processes in the two simulation experiments of section 4.3. Left: *VAR(1)* fitted under a parametric *VARMA(1,1)* model; Right: *VAR(2)* fitted under a parametric *VARMA(2,2)* model

4.4 Comparisons with previous criteria

As seen in table 11, the adjusted residual trace $\mathrm{tr}(\widehat{\mathbf{R}}_k)/\sqrt{m}$, and the corresponding modified version $\mathrm{tr}(\widehat{\mathbf{S}}_k)/\sqrt{m}$ are virtually identical for large enough values of the lag k. Thus, it seems adequate to consider a truncated version $\{\widehat{Z}^m_{n,M}(u) : 0 \leq u \leq 1\}$ of the modified process $\{\widehat{Z}^m_n(u) : 0 \leq u \leq 1\}$, where the random function $\widehat{Z}^m_n(u)$ is replaced by

$$\widehat{Z}^m_{n,M}(u) = \frac{\sqrt{2}}{\pi}\sqrt{n}\left[\sum_{k=p+q+1}^{M} \frac{\mathrm{tr}(\widehat{\mathbf{S}}_k)}{\sqrt{m}}\frac{\sin(K\pi u)}{K} + \sum_{k=M+1}^{n-(P+1)} \frac{\mathrm{tr}(\widehat{\mathbf{R}}_k)}{\sqrt{m}}\frac{\sin(K\pi u)}{K}\right] , \qquad (69)$$

for an adequately chosen value of $M \geq p + q + 1$. This idea was also explored in the univariate case by Ubierna and Velilla (2007, section 4.1).

As such, goodness-of-fit functionals of the form

$$CVT = \int_0^1 [\widehat{Z}^m_{n,M}(u)]^2 du = \frac{n}{m\,\pi^2}[\sum_{k=p+q+1}^{M} \frac{[\mathrm{tr}(\widehat{\mathbf{S}}_k)]^2}{K^2} + \sum_{k=M+1}^{n-(P+1)} \frac{[\mathrm{tr}(\widehat{\mathbf{R}}_k)]^2}{K^2}] \; ; \qquad (70)$$

and

$$KST = \sup_{1\leq j\leq n} |\widehat{Z}^m_{n,M}(j/n)| \; , \qquad (71)$$

can be compared to the standard criteria of Hosking (1980),

$$\widehat{Q}^m_H = n\sum_{k=1}^{M}\mathrm{tr}(\widehat{\mathbf{C}}'_k\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{C}}_k\widehat{\mathbf{\Sigma}}^{-1}) \; ;$$

and Li and McLeod (1981),

$$\widehat{Q}^m_{LM} = n\sum_{k=1}^{M}\mathrm{tr}(\widehat{\mathbf{C}}'_k\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{C}}_k\widehat{\mathbf{\Sigma}}^{-1}) + \frac{m^2M(M+1)}{2n} \; .$$

For a nominal level α, the rejection regions for CVT and KST are based on the critical points of the corresponding functionals of the Brownian bridge. Regions associated to $\widehat{Q}^m_H$ and $\widehat{Q}^m_{LM}$ use the chi−square quantile $\chi_{m^2[M-(p+q)],\alpha}$. Comparisons are performed now, both in size and power, using simulation techniques.

4.4.1 Size

In principle, in $\widehat{Q}^m_H$ and $\widehat{Q}^m_{LM}$ the value of M is taken of the order $O(\sqrt{n})$. It is however of interest to study the dependence on M of the size of the four methods above. Two

models are considered. First, a *VAR(1)* $\mathbf{X}_t = \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t$ in which

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.5603 & 0.5361 \\ -0.0366 & 0.6303 \end{pmatrix} . \tag{72}$$

The eigenvalues of the matrix $\mathbf{\Phi}_1$ in (72) are $\delta_j = 0.5953 \pm 0.1356\, i$, $j = 1, 2$, so that $|\delta_1| = |\delta_2| = 0.6106 < 1$. The associated eigenvectors are $\boldsymbol{\gamma}_1 = (0.9675, 0.0632 + 0.2447\, i)'$; $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (0.9675, 0.0632 - 0.2447\, i)'$. An additional *VMA(1)* process of the form $\mathbf{X}_t = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1 \boldsymbol{\varepsilon}_{t-1}$ is analyzed, with

$$\mathbf{\Theta}_1 = \begin{pmatrix} 0.0589 & 0.3047 \\ -0.0093 & 0.1212 \end{pmatrix} . \tag{73}$$

The eigenvalues of $\mathbf{\Theta}_1$ in (73) are $\delta_j = 0.0901 \pm 0.0433\, i$, $j = 1, 2$, so that $|\delta_1| = |\delta_2| = 0.0999 < 1$. The eigenvectors are $\boldsymbol{\gamma}_1 = (0.9850, 0.1007 + 0.1400\, i)'$; $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (0.9850, 0.1007 - 0.1400\, i)'$. In both cases, the covariance matrix $\mathbf{\Sigma}$ of the errors is taken as in (54). For each model, $N = 1000$ independent replicas are generated. The sample size considered for the *VAR(1)* model is $n = 250$; and $n = 200$ for the *VMA(1)*. In both cases, the nominal level is $\alpha = .05$.

For values of $2 \leq M \leq 40$, figures 8 and 10 display the resulting empirical sizes for the *VAR(1)* and *VMA(1)* cases, respectively. Both plots indicate that the size of both *CVT* in (70) and *KST* in (71) is relatively stable with respect the value of M. As it can be seen in these figures, this typically falls inside the horizontal bands $.05 \pm 1.96\sqrt{0.05 \times 0.95/1000}$. In contrast, the size of Hosking (1980) decreases. In turn, that of Li and McLeod (1981) is much less stable, being above the nominal level $\alpha = 0.05$ in the *VMA(1)* setting.

4.4.2 Power

For making comparisons in power, two models are considered. First, the parametric *VARMA(1,1)* process $\mathbf{X}_t - \mathbf{\Phi}_1 \mathbf{X}_{t-1} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1(\beta) \boldsymbol{\varepsilon}_{t-1}$ defined by the setting (65)−(66)−(54). A *VAR(1)* is tested at the nominal level $\alpha = .05$. For each value of $0 \leq \beta < 1$, $N = 1000$ independent data samples of length $n = 200$ are generated. The value of the lag M is taken as integer part of $\sqrt{n}$. Hence, $M = 14$. Results are displayed at the left part of figure 10.

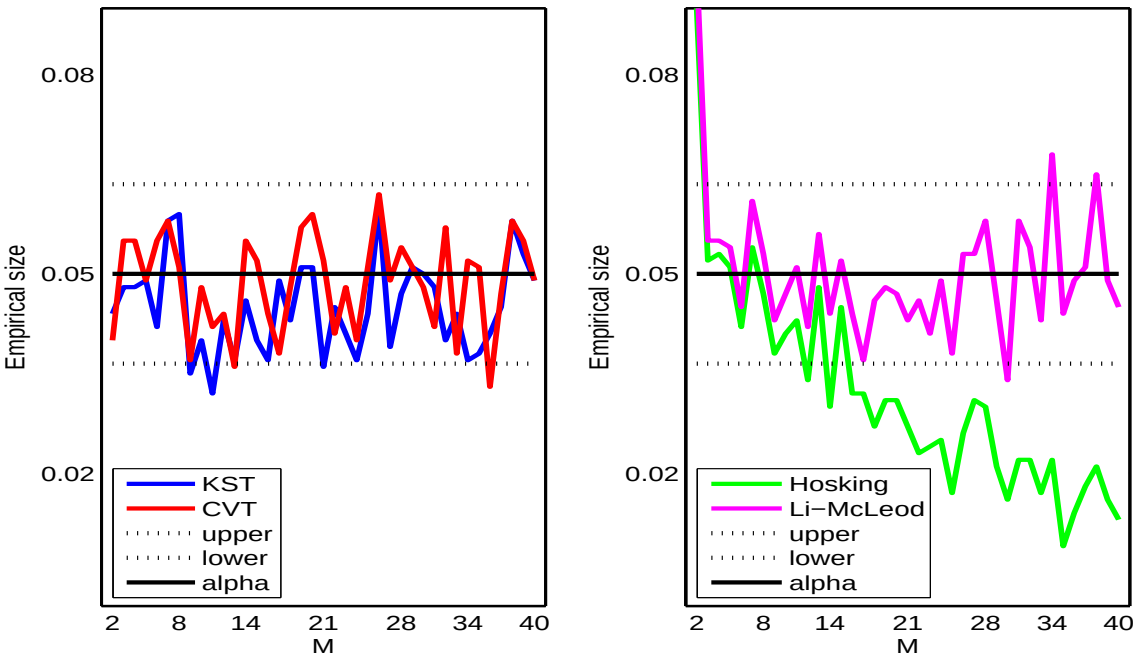

FIG. 8: Comparison of empirical sizes for different values of the lag M for the *VAR(1)* model of section 4.4.1

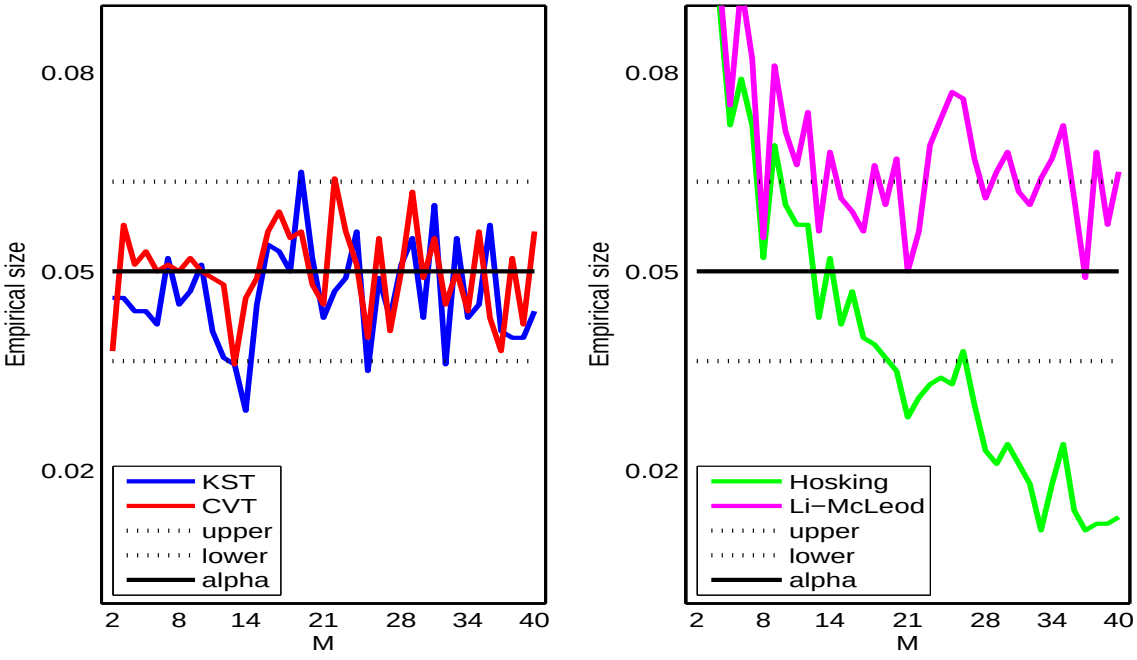

FIG. 9: Comparison of empirical sizes for different values of the lag M for the *VMA(1)* model of section 4.4.1

Alternatively, a *VARMA(1,1)* $\mathbf{X}_t - \mathbf{\Phi}_1(\beta)\mathbf{X}_{t-1} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1\boldsymbol{\varepsilon}_{t-1}$ is considered, where $\mathbf{\Theta}_1$

is as in (53); $\mathbf{\Sigma}$ as in (54); and $\mathbf{\Phi}_1(\beta) = \beta\mathbf{\Phi}_1$, $0 \leq \beta < 1$, where

$$\mathbf{\Phi}_1 = \begin{pmatrix} 0.9177 & 0.8780 \\ -0.0599 & 1.0324 \end{pmatrix} . \tag{74}$$

The eigenvalues of $\mathbf{\Phi}_1$ in (74) are $\delta_j = 0.9750 \pm 0.2221\, i$, $j = 1, 2$, so that $|\delta_1| = |\delta_2| = 1.0000$. The eigenvectors are $\boldsymbol{\gamma}_1 = (0.9675, 0.0632 + 0.2447\, i)'$, and $\boldsymbol{\gamma}_2 = \overline{\boldsymbol{\gamma}}_1 = (0.9675, 0.0632 - 0.2447\, i)'$. A *VMA(1)* is now tested at level $\alpha = .05$. Choices for N, n, and M are as before. Results are given at the right part of figure 10.

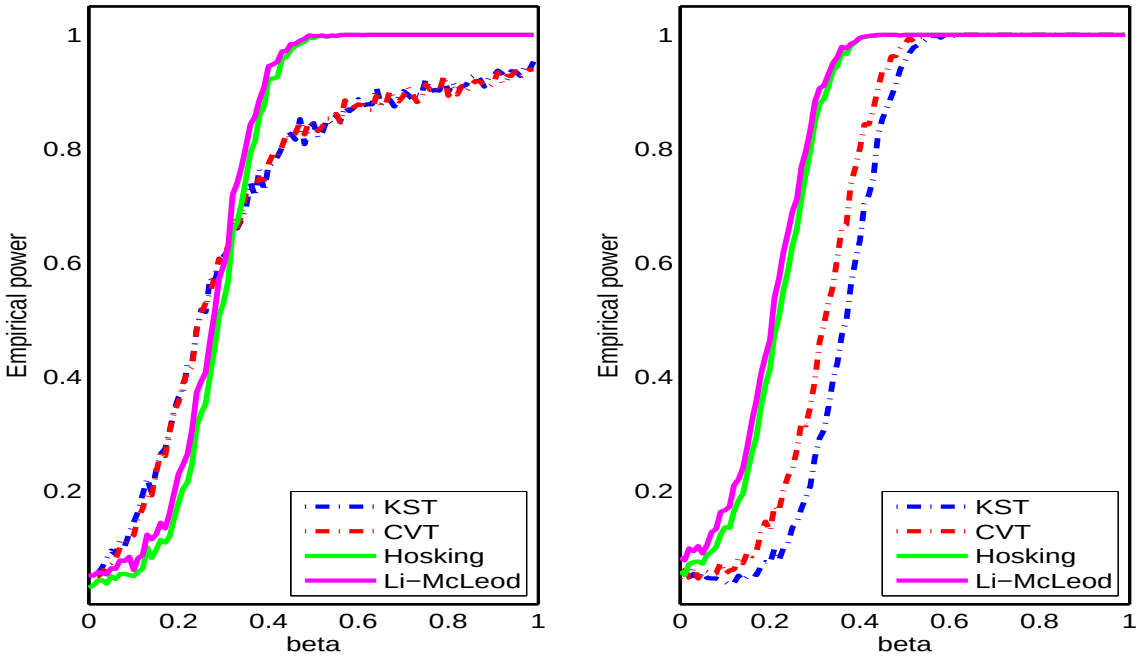

FIG. 10: Comparison of empirical powers for a fixed value of the lag M in the first two simulation experiments of section 4.4.2. Left: *VAR(1)* fitted under a parametric *VARMA(1,1)* model; Right: *VMA(1)* fitted under a parametric *VARMA(1,1)* model

According to figure 10, our methods are locally more powerful than those by Hosking (1980) and Li and McLeod (1981) when a *VAR(1)* is tested. In turn, the latter procedures clearly outperform *KST* and *CVT* when the postulated null model is a *VMA(1)*. In this case, all the power functions tend to 1 when $\beta \to 1$.

Finally, the parametric *VARMA(2,2)* model $\mathbf{X}_t - \mathbf{\Phi}_1\mathbf{X}_{t-1} - \mathbf{\Phi}_2\mathbf{X}_{t-2} = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1(\beta)\boldsymbol{\varepsilon}_{t-1} + \mathbf{\Theta}_2(\beta)\boldsymbol{\varepsilon}_{t-2}$, $0 \leq \beta < 1$, of section 4.3.2 is revisited. A *VAR(2)* is now postulated and fitted. Choices for the tuning constants N, n, M, and α are as above. Results are in

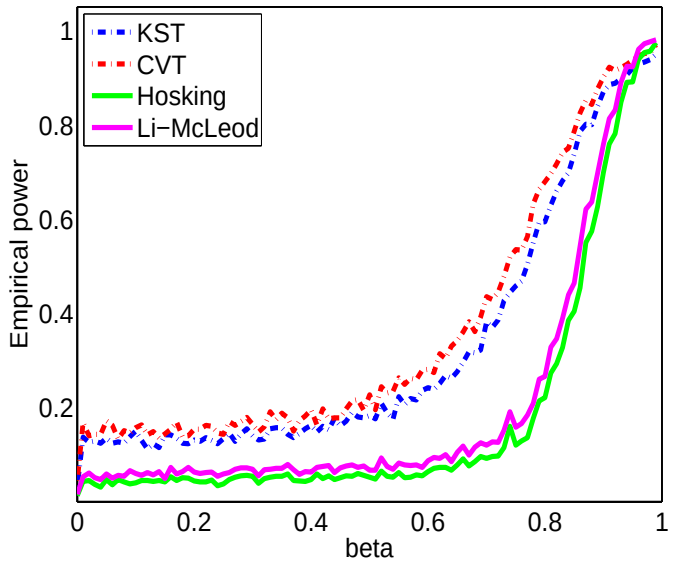

FIG. 11: Comparison of empirical powers for a fixed value of the lag M in the third simulation experiment of section 4.4.2. A *VAR(2)* is fitted under a parametric *VARMA(2,2)* model

figure 11. As seen there, the empirical power of our procedures is above those of the standard methods by Hosking (1980) and Li and McLeod (1981).

5 CONCLUSIONS

This research introduces a new goodness-of-fit process for *VARMA(p,q)* models. We extend to a multivariate setting the univariate goodness-of-fit process studied by Durlauf (1991) and Anderson (1993). We study weak convergence of this new method, whose application does not depend on the choice of a particular lag, and that uses a distribution that is free of unknown parameters. Simulations and comparisons with previous methods are also given. As a conclusion of this empirical work, our procedure seems to be sensitive enough to detect lack of fit.

6 References

Anderson, T. W., 1955. The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. Proceedings of the American Mathematical Society 6 (2), pp. 170–176.

Anderson, T. W., 1993. Goodness of fit tests for spectral distributions. The Annals of Statistics 21 (2), pp. 830–847.

Chitturi, R. V., 1974. Distribution of residual autocorrelations in multiple autoregressive schemes. Journal of the American Statistical Association 69 (348), 928–934.

Dunsmuir, W., Hannan, E. J., 1976. Vector linear time series models. Advances in Applied Probability 8 (2), pp. 339–364.

Durbin, J., 1975. Tests of model specification based on residuals. Vol. A Survey of Statistical Design and Linear Models. North-Holland, Amsterdam.

Durlauf, S. N., 1991. Spectral based testing of the martingale hypothesis. Journal of Econometrics 50 (3), 355 – 376.

Hannan, E. J., 1969. The identification of vector mixed autoregressive-moving average systems. Biometrika 56 (1), pp. 223–225.

Hosking, J. R. M., 1980. The multivariate portmanteau statistic. Journal of the American Statistical Association 75 (371), pp. 602–608.

Li, W. K., McLeod, A. I., 1981. Distribution of the residual autocorrelations in multivariate arma time series models. Journal of the Royal Statistical Society. Series B (Methodological) 43 (2), pp. 231–239.

Lütkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. Springer.

Ubierna, A., Velilla, S., 2007. A goodness-of-fit process for arma models based on a modified residual autocorrelation sequence. Journal of Statistical Planning and Inference 137 (9), 2903 – 2919.

Velilla, S., Nguyen, H., 2011. A basic goodness-of-fit process for varma(p,q) models. Statistics and econometrics series 09, Carlos III University of Madrid.